Function-on-Function Regression with Public Health Applications

Function-on-Function Regression with Public Health Applications

A dissertation presented

by

Mark John Meyer

to

The Department of Biostatistics

in partial fulfillment of the requirements
for the degree of
Doctor of Philosophy
in the subject of
Biostatistics

Harvard University
Cambridge, Massachusetts

May 2014

Dissertation Advisor: Professor Brent A. Coull Mark John Meyer

Function-on-Function Regression with Public Health Applications

Abstract

Medical research currently involves the collection of large and complex data. One such type of data is functional data where the unit of measurement is a curve measured over a grid. Functional data comes in a variety of forms depending on the nature of the research. Novel methodologies are required to accommodate this growing volume of functional data alongside new testing procedures to provide valid inferences. In this dissertation, I propose three novel methods to accommodate a variety of questions involving functional data of multiple forms. I consider three novel methods: (1) a function-on-function regression for Gaussian data; (2) a historical functional linear models for repeated measures; and (3) a generalized functional outcome regression for ordinal data. For each method, I discuss the existing shortcomings of the literature and demonstrate how my method fills those gaps. The abilities of each method are demonstrated via simulation and data application.

Contents

List of Figures

List of Tables

In memory of Gang Zheng.

Acknowledgments

I would like to first thank my parents, Kent and Deb, and my sister, Serenity, whose love and support over the years brought me through tough times. I'd like especially thank my advisor, Brent A. Coull, who has served as a wonderful mentor over the last five years and without whom none of this would be possible. Next, I'd like to thank my previous mentors, Monica Jackson, Gang Zheng, and Elizabeth (Betty) J. Malloy, without whom I would not be where I am. In particular, I'd like to give additional thanks to Betty who inspired me to pursue statistical research, encouraged me to apply to graduate school, served on my committee, and continues to be a mentor, collaborator, and friend. To the remaining members of my committee, Christoph Lange and Craig P. Hersh, thank you for your contributions to this dissertation and the knowledge you've imparted. I'd like to give special thanks to Matt Cefalu, Natalie Exner, Lauren Kunz, Shira Mitchell, and Elizabeth Smoot who all have provided an abundance of moral support over the past few years. Finally, I'd like to thank my friends and extended family who have supported me over these years.

1. Bayesian Function-on-Function Regression for Multi-Level Functional Data

Mark J. Meyer[1], Brent A. Coull[1], Francesco Versace[2], Paul Cinciripini[2], and Jeffrey S. Morris[2]

[1]Department of Biostatistics, Harvard School of Public Health
[2]The University of Texas M.D. Anderson Cancer Center

1.1 Introduction

Medical and public health research increasingly involves the collection of complex and high dimensional data. In particular, functional data—where the unit of observation is a curve or set of curves that are finely sampled over a grid—is frequently obtained (Ramsay and Silverman, 2005). Moreover, researchers often sample multiple curves per subject which yields repeated functional measures. A common question is how to analyze the relationship between two functional variables. While the field of functional data analysis (FDA) has progressed considerably in recent years, gaps remain in the literature with regards to function-on-function regression where both the predictor and outcome are functional.

Regression in FDA can be classified into three broad sub-classes: scalar-on-function, function-on-scalar, and function-on-function. Classical functional regression, on which a large literature exists, involves scalar-on-function regression where the outcome is scalar and the predictor is functional, with functional regression coefficients. See for instance Ramsay and Dalzell (1991), Cardot, Ferraty, and Sarda (1999), Reiss and Ogden (2007), Malloy et al. (2010), Goldsmith et al. (2011), McLean et al. (2012), Gertheiss, Maity, and Staicu (2013), and references therein. Function-on-scalar regression, also heavily investigated in the literature, involves regressing a functional predictor on to a set of scalar covariates, each of which has a functional regression coefficient. See for instance Brumback and Rice (1998), Morris and Carroll (2006), Reiss, Huang, and Mennes (2010), Staicu et al. (2011), Chen and Müller (2012), Goldsmith, Greven, and Crainiceanu (2013), and references therein.

In contrast, the literature addressing function-on-function regression, with functional outcome, functional predictor, and a coefficient surface, is rather sparse. Much of it is dedicated to the historical functional linear model (HFLM), as described by Malfait and Ramsay (2003) and further examined by Harezlak et al. (2007) and Kim, Şentürk, and Li (2011). The primary assumption in an HFLM is that the association between curves is

uni-directional, which leads to a upper triangular regression surface. That is, for functions of time, an association between the predictor at any given time-point can only occur with the outcome at subsequent times. Function-on-function regression allowing for bi-directional associations—that is, with unconstrained regression coefficient surfaces—is explored by Yao, Müller, and Wang (2005) and Müller and Yao (2008).

There are some recent technical reports on the topic from one research group, Ivanescu et al. (2012), Scheipl and Greven (2012), Scheipl, Staicu, and Greven (2014), that discuss a penalized spline approach, identifiability issues, and function-on-function regression in Functional Additive Mixed Models, respectively. One major limitation to both Scheipl and Greven (2012) and Scheipl, Staicu, and Greven (2014) is the assumption of iid errors which might be unrealistic for functional data. Ivanescu et al. (2012) do allow for correlated errors, but only present results assuming iid errors. Additionally, neither Scheipl and Greven (2012) nor Ivanescu et al. (2012) account for correlation induced by multiple measurements on the same subjects and while Scheipl, Staicu, and Greven (2014) do incorporate a random functional effects, they cannot incorporate correlation between different random effects. Ivanescu et al. (2012) and Scheipl, Staicu, and Greven (2014) address inferential procedures, relying on 95% point-wise confidence intervals (PWCI) to determine significance. Neither approach, however, makes any adjustment for the multitude comparisons.

To motivate our development of the function-on-function setting, we examine data from a smoking cessation trial, conducted in the Department of Behavior Sciences at the University of Texas M. D. Anderson Cancer Center (Cinciripini et al., 2013). For a subset of participants, researchers obtained Event Related Potentials (ERPs) at baseline during the presentation of a series of images depicting neutral, positive, negative, and cigarette-related contents. ERPs were collected using a 129 channel Geodesic Sensor Net. Finely sampled curves were produced over the course of 900 ms (100 ms prior to picture presentation and 800 ms after). Electrical potentials every 4ms were collected from 129 electrodes distributed on the surface of the scalp resulting in 225 measurements for each electrode.

While many analyses are of interest for these data, in this paper we focus on characterizing the time-varying relationship between ERP outputs from pairs of electrodes.

In this paper, we propose a general function-on-function regression modeling framework that can accommodate this type of multilevel functional data. The model is flexible enough to incorporate a variety of basis expansions including such common approaches as principal components, spline-based and wavelet-based functional representations. Our approach not only allows for correlation between functions through random effect functions, but also allows heteroscedasticity and within-function correlation in the residual error functions. While the approach can be applied generally for any number of functional predictors and arbitrary interactions with other discrete and continuous predictors, we present specific model formulations for both a single functional predictor of interest as well an interaction of a discrete factor with a functional predictor which results in separate function-on-function regressions for each discrete factor. For inference, we propose three approaches. First we extend the Bayesian False Discovery Rate (BFDR) procedure proposed by Morris et al. (2008) to the function-on-function setting. Second, we generate joint credible bands as in Ruppert, Wand, and Carroll (2003). Next we generate two novel Bayesian summaries: (1) Simultaneous Band Scores (SimBaS), a functional measure we introduce that summarizes for each position in the regression surface the smallest α for which the $100(1-\alpha)\%$ joint credible bands exclude zero at that position, and (2) Global Bayesian P-Values (GBPV), which can be interpreted as a type of Bayesian p-value corresponding to a global functional null hypothesis of no relationship between the functions. These summaries are of general interest and can be used in other functional regression settings.

Section 1.2 develops a simple version of our proposed function-on-function mixed model, presents a more general model, and describes our basis function modeling strategy. Section 1.3 details the BFDR, SimBaS, and GBPV inference procedures. In Section 1.4 we present the results of a simulation assessing model fit and the BFDR, SimBaS, and GBPV procedures. Section 1.5 presents the results obtained by applying the proposed methods

to the ERP data, and Section 1.6 contains further discussion.

1.2 Function-on-Function Regression Model for Multi-Level Functional Data

Here we introduce the function-on-function model we will use to regress one function $y(t), t \in \mathcal{T}$ on another $x(v), v \in \mathcal{V}$. First we consider a simple case with a single functional predictor and repeated measures of $\{y(t), x(v)\}$ pairs for each subject, and then in Section 1.2.3 we describe more complex models that can be handled by our approach.

Individual subjects are denoted as $i = 1, \ldots, n$. Let $c = 1, \ldots, C_i$ index repeated pairs of curves observed on subject i. Then for subject i, curve set c, we observe $x_{ic}(v)$ and $y_{ic}(t)$, $\{y_{ic}(t), x_{ic}(v) : t \in \mathcal{T}, v \in \mathcal{V}\}$,

$$y_{ic}(t) = \alpha(t) + \int_{v \in \mathcal{V}} x_{ic}(v)\beta(v,t)dv + U_i(t) + E_{ic}(t). \tag{1.1}$$

We assume observation-specific and subject-specific Gaussian process errors $E_{ic}(t) \sim \mathcal{GP}(0, \Sigma_E)$ and $U_i(t) \sim \mathcal{GP}(0, \Sigma_U)$. The integration over the entire support of v allows the exposure-response relationship to move in either direction, i.e. we do not assume the timing of an effect of x on y occurs in one direction or the other. That relationship is characterized by the surface $\beta(v,t)$.

In this paper, our focus is on functional data sampled on a common fine grid. Here, we consider a discretized version of Model (1.1). Let $y_{ic}(\cdot)$ be finely sampled on a grid $\mathbf{t} = [t_1 \ \cdots \ t_T]$ of length T. Similarly, $x_{ic}(\cdot)$ is observed on a grid $\mathbf{v} = [v_1 \ \cdots \ v_V]$ of length V. We can then define the row vectors $\mathbf{y}_{ic} = [y_{ic}(t_1) \ \cdots \ y_{ic}(t_T)]$ and $\mathbf{x}_{ic} = [x_{ic}(v_1) \ \cdots \ x_{ic}(v_V)]$ and express Model 1.1 in the discrete form

$$\mathbf{y}_{ic} = \mathbf{x}_{ic}\boldsymbol{\beta} + \mathbf{u}_i + \mathbf{e}_{ic} \tag{1.2}$$

where $\mathbf{y}_{ic}$, $\mathbf{u}_i$, and $\mathbf{e}_{ic}$ are $1 \times T$, $\mathbf{x}_{ic}$ is $1 \times V$, and $\boldsymbol{\beta}$ is the $V \times T$ matrix of coefficients. Note then that $\mathbf{e}_{ic} \sim \mathcal{N}(0, \Sigma_E)$ and $\mathbf{u}_i \sim \mathcal{N}(0, \Sigma_U)$. In practice, we center and scale both $y_{ic}(t)$ and $x_{ic}(v)$ and thus

Now let N be the total number of observed response curves. Stacking the row vectors by subject, $\mathbf{Y}$ and $\mathbf{X}$ represent the $N \times T$ and $N \times V$ matrices of observed curves. Further, $\mathbf{Z}$ is the $N \times n$ random effects design matrix. Our discretized model for all subjects is then

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}U + \mathbf{E} \tag{1.3}$$

where $\boldsymbol{\beta}$ is as defined in Model (1.2), U is the $n \times T$ matrix of subject specific random effect functions on the grid, and $\mathbf{E}$ is the $N \times T$ matrix of model errors, interpretable as residual curve-to-curve deviations.

Because of the functional nature of the data, we do not directly fit Model (1.3). Instead, we represent the curves using some basis function expansion and apply basis transformations to $y(t)$ and $x(v)$ prior to model fitting. This basis function transform approach has numerous advantages, including dimension reduction, more efficient computation, and borrowing of strength across observations of the curves. Previous work in the functional regression context has used a variety of basis functions including kernels, splines, wavelets, and functional Principal Components (fPC). We will begin by presenting a generalized basis expansion for our model to demonstrate how multiple candidate transformations can be used in our model. Then we will present the rest of the modeling details using specific basis functions chosen for our simulation and data analysis, with the understanding that it can be adapted for use with other basis functions.

1.2.1 General Basis Transform Modeling Approach

Here we describe our general basis function transform approach for fitting the function-on-function regression models, which involves projecting both the functional responses and predictors into a chosen basis space, fitting the model in the basis space, and then transforming the results back to the original function space for interpretation and inference.

Let $y_{ic}(t) = \sum_{j=1}^{T^*} y^*_{icj}\xi_j(t)$ and $x_{ic}(v) = \sum_{j=1}^{V^*} x^*_{icj}\phi_j(v)$ be some chosen truncated basis

expansion for the functional responses and predictors, respectively. Potential choices include wavelets, B-splines, kernels, Fourier bases, principal components, or independent components. Let ξ be a matrix of size $T^* \times T$ containing the basis functions on the discrete grid $\mathbf{t}$ with element (i, j) given by $\xi_i(t_j)$, and likewise let ϕ be a $V^* \times V$ matrix containing the basis functions for $x(v)$ evaluated on the grid $\mathbf{v}$. Considering the discretely sampled functions in matrix form, we can write the basis expansion as $\mathbf{Y} = \mathbf{Y}^*\xi$ and $\mathbf{X} = \mathbf{X}^*\phi$, with $\mathbf{Y}^*$ and $\mathbf{X}^*$ being $N \times T^*$ and $N \times V^*$ matrices, respectively, containing the basis coefficients for the observed functions. Here we assume that ϕ and ξ are of full row rank, possibly but not necessarily orthogonal, so rank$(\phi) = V^*$, rank$(\xi) = T^*$ and $\phi\phi'$ and $\xi\xi'$ are invertible matrices of size $V^* \times V^*$ and $T^* \times T^*$, respectively.

Replacing each functional quantity in Model (1.3) with its basis expansion, we have

$$\mathbf{Y}^*\xi = \mathbf{X}^*\phi\phi'\boldsymbol{\beta}^*\xi + \mathbf{Z}U^*\xi + \mathbf{E}^*\xi, \tag{1.4}$$

where $\boldsymbol{\beta}^*$ is $V^* \times T^*$, U^* is $n \times T^*$, and $\mathbf{E}^*$ is $N \times T^*$, representing quantities of Model (1.3) in the transformed basis space. When ϕ is orthogonal so that $\phi\phi' = I_{V^*}$, if we multiply each side of (1.4) by $\xi^- = \xi'(\xi\xi')^{-1}$, then we arrive at the *basis space model*

$$\mathbf{Y}^* = \mathbf{X}^*\boldsymbol{\beta}^* + \mathbf{Z}U^* + \mathbf{E}^*. \tag{1.5}$$

When ϕ is not orthogonal, we instead replace $\boldsymbol{\beta}^*$ in Model (1.5) with $\boldsymbol{\beta}^\dagger = \phi\phi'\boldsymbol{\beta}^*$. Thus, we can fit this basis space model after first transforming the functional responses and predictors to their respective basis spaces, $\mathbf{Y}^* = \mathbf{Y}\xi^-$ and $\mathbf{X}^* = \mathbf{X}\phi^-$, with $\phi^- = \phi'(\phi\phi')^{-1}$, and then after fitting the model, transform back to the original function space to obtain estimates and inference for $\boldsymbol{\beta} = \phi'\boldsymbol{\beta}^*\xi$ when ϕ is orthogonal, $\boldsymbol{\beta} = \phi^-\boldsymbol{\beta}^\dagger\xi$ otherwise. Note that for some choices of basis functions, fast transform algorithms can be used in lieu of matrix multiplication to compute the basis functions or transform back to the original space, e.g., discrete wavelet transform (DWT) for wavelets, discrete Fourier transform (DFT) for Fourier bases, and fast algorithms for computing independent components (Hyvarinen et al., 2001).

We take a Bayesian approach to fit Model (1.5), using an Markov Chain Monte Carlo (MCMC) procedure to sample from the posterior distributions using appropriate prior distributions for each model parameter. The specifics of the sampler may vary slightly depending on choice of basis function, and will be described in Web Appendix A. This formulation allows a variety of possible basis functions to be used for the outcome and predictor, including variations of wavelets, principal components, Fourier series, and splines, each of which corresponds to different choices of ξ and ϕ. For example, for wavelets ξ and ϕ are inverse discrete wavelet transform (IDWT) matrices, for principal components they are the eigenvectors, possibly rescaled by the eigenvalues, for Fourier series they are the Inverse Discrete Fourier Transform (IDFT) matrices, and for splines they can be constructed based on B-splines or orthogonalized B-spline design matrices. Note that the same basis transform does not need to be used for both $y(t)$ and $x(v)$. In this paper, we use wavelet bases to represent the functional form of $y(t)$, and for $x(v)$, we use a composite strategy involving wavelets followed by principal components that we refer to as wPC, which is similar to strategies used by Johnstone and Lu (2009) and Røislien and Winje (2012).

1.2.2 Model Formulation

Here, we present our modeling details using wavelets for $y(t)$ and wPC for $x(v)$. First, we transform the functions to the wavelet space by applying the $O(T)$ DWT to each row of $\mathbf{Y}$ and $\mathbf{X}$, which can be represented as

$$\mathbf{y}_{ic} \xrightarrow{\text{DWT}} \mathbf{y}_{ic}^{W} = \{y_{ic,jk}^{W}\} \text{ and } \mathbf{x}_{ic} \xrightarrow{\text{DWT}} \mathbf{x}_{ic}^{W} = \{x_{ic,s\ell}^{W}\}.$$

Wavelets are multi-resolution bases that are double-indexed by scale and location. The scales are $j = 1, \ldots, J^y$ and $s = 1, \ldots, S^x$ and locations $k = 1, \ldots, K_j^y$ and $\ell = 1, \ldots, L_s^x$ for $\mathbf{Y}$ and $\mathbf{X}$, respectively. The dimension of $\mathbf{y}_{ic}^{W}$ is $1 \times T^W$ where $T^W = \sum_{j=1}^{J^y} k_j^y$. Similarly, $\mathbf{x}_{ic}^{W}$ has dimensions $1 \times V^W$ where $V^W = \sum_{s=1}^{S^x} \ell_s^x$. If T and V are powers of two, this decomposition will result in $T^W = T$ and $V^W = V$ wavelet coefficients, and otherwise

padding is done according to some chosen boundary condition (e.g. periodic, reflection, and padding with zeros), in which case T^W and V^W are not exactly equal to but are of the same order as T and V. We discuss choice of padding further in our simulation study in Section 1.4.

Wavelets tend to provide sparse representations for many functions, so one can achieve data compression by eliminating wavelet coefficients that are negligible in magnitude for all curves. Wavelet thresholding has been widely used for compression and denoising of individual functions, and Morris et al. (2011) introduced a *joint compression* approach for the multiple function setting that finds a minimal subset of wavelet coefficients that jointly preserves $100\alpha\%$ of the total energy for *all* functions in a set. Let T^{W^*} and V^{W^*} represent the total number of coefficients left after such joint compression.

We can write these wavelet basis expansions in matrix form as $\mathbf{Y} = \mathbf{Y}^W W_y$ and $\mathbf{X} = \mathbf{X}^W W_x$, where W_y and W_x are $T^{W^*} \times T$ and $V^{W^*} \times V$ matrices, respectively, containing the retained wavelet basis functions evaluated on the T and V grids. Given orthogonal wavelets, we can also represent the DWT in matrix form as $\mathbf{Y}^W = \mathbf{Y}W_y'$ and $\mathbf{X}^W = \mathbf{X}W_x'$, or if non-orthogonal they can be represented $\mathbf{Y}^W = \mathbf{Y}W_y^-$ and $\mathbf{X}^W = \mathbf{X}W_x^-$. Thus, in the notation of Section 1.2.1, if we use wavelet transforms with joint compression for both $y(t)$ and $x(v)$, then we effectively define $\xi = W_y$ and $\phi = W_x$, with $\xi^- = W_y'$ and $\phi^- = W_x'$, $\mathbf{Y}^* = \mathbf{Y}^W$ and $\mathbf{X}^* = \mathbf{X}^W$, and $T^* = T^{W^*}$ and $V^* = V^{W^*}$.

In our model, calculations are linear in T^* but quadratic in V^*, so dimension reduction in $\mathbf{X}^*$ has especially important computational benefits. While the joint compression provides some dimension reduction, use of Principal Components Analysis (PCA) can provide additional dimension reduction. In particular, consider performing the singular value decomposition of $\mathbf{X}^W = \mathbf{X}W_x' = Q\Sigma P'$. Noting that $\mathbf{X}^W$ is $N \times V^{W^*}$, we see that Q, the matrix of left singular vectors, is $N \times V^{W^*}$ and both Σ, the diagonal matrix of singular values, and P, the matrix of right singular vectors, are $V^{W^*} \times V^{W^*}$. Supposing we keep $V^{svd} \ll V^{W^*}$ principal components, we can compute the wavelet-space PC scores $\mathbf{X}^* =$

$\mathbf{X}^W P_{svd}$, where P_{svd} is a $V^{W*} \times V^{svd}$ matrix computed from the leading V^{svd} rows of P. Using the notation of Section 1.2.1, this composite basis function strategy is equivalent to computing $\mathbf{X}^* = \mathbf{X}\phi^-$ with ***composite transform*** $\phi^- = W_x' P_{svd}$ and inverse transform $\phi = P_{svd}' W_x$ of dimension $V^* = V^{svd}$. Note that one could simply define ϕ to be the eigenvectors of a direct SVD on $\mathbf{X}$, but this composite wPC approach has advantages in that the joint compression in the wavelet space (1) reduces the dimensionality of $\mathbf{X}$ to speed up calculation of the SVD, (2) performs some denoising of the functions in $\mathbf{X}$ before calculation of the SVD, and (3) borrows strength locally within the function, thus accounting for the functional nature of the data.

Thus, after transforming the data, recall our basis space model (1.5) is given by $\mathbf{Y}^* = \mathbf{X}^*\boldsymbol{\beta}^* + \mathbf{Z}U^* + \mathbf{E}^*$. Consistent with previous work (Morris and Carroll (2006), Morris et al. (2008), Zhu, Brown, and Morris (2011), among others), we assume independence in the wavelet space. That is, for the subject specific version of Model (1.5), $\mathbf{y}_{ic}^* = \mathbf{x}_{ic}^*\boldsymbol{\beta}^* + \mathbf{u}_i^* + \mathbf{e}_{ic}^*$, we assume $\mathbf{e}_{ic}^* \sim \mathcal{N}(0, \Sigma_e^*)$ where Σ_e^* is a diagonal matrix with elements varying by j, k, $\Sigma_e^* = \left\{\sigma_{e(j,k)}^2\right\}$, and equivalently $\mathbf{u}_i^* \sim \mathcal{N}(0, \Sigma_u^*)$ where $\Sigma_U^* = \text{diag}\left\{\sigma_{U(j,k)}^2\right\}$. The induced within-function covariances in the data space are given by $\Sigma_e = \xi'\Sigma_e^*\xi$ and $\Sigma_u = \xi'\Sigma_u^*\xi$, which with wavelets accommodates a broad class of covariances allowing heteroscedasticity and differing degrees of autocorrelation, and thus different degrees of borrowing of strength, in different regions of the function (Morris and Carroll, 2006; Morris, et al. 2008; Morris, et al. 2011). When other basis functions are used, one must consider whether the class of induced covariance structures from basis space independence is sufficiently flexible to capture the key functional features, with other parsimonious alternatives possible, for example serial correlation across neighboring basis coefficients.

The basis space independence assumption allows us to split Model (1.5) into a series of T^* separate models for each basis coefficient in the y-space, double-indexed by (j, k), giving $\mathbf{y}_{(j,k)}^* = \mathbf{X}^*\beta_{(j,k)}^* + \mathbf{Z}\mathbf{u}_{(j,k)}^* + \mathbf{e}_{(j,k)}^*$, where $\mathbf{y}_{(j,k)}^*$ and $\mathbf{e}_{(j,k)}^*$ are $N \times 1$, $\beta_{(j,k)}^*$ is $V^* \times 1$, and $\mathbf{u}_{(j,k)}^*$ is $n \times 1$. $\mathbf{X}^*$ and $\mathbf{Z}$ are as previously defined. This separability allows computational scalability to extremely large T, as calculations are linear in T^*, sparse basis functions

frequently yield $T^* \ll T$, and when cluster computing resources are available, allows parallel computing across (j, k). For prior specification, we assume vague proper priors on the variance components and a spike-and-slab prior similar to that found in Morris and Carroll (2006), Malloy et al. (2010), and others (see Web Appendix A for details). Posterior samples are then generated for $\boldsymbol{\beta}^*$ and projected back into the data-space using $\boldsymbol{\beta} = \phi^- \boldsymbol{\beta}^* \xi$, where recall for our example $\phi^- = \phi' = P'_{svd} W_x$ and $\xi = W_y$. These posterior samples are used to perform Bayesian inference on $\boldsymbol{\beta}$, as detailed in Section 1.3.

1.2.3 More Complex Function-on-Function Mixed Models

The simple function-on-function regression Model (1.1) is a special case of a general function-on-function mixed model that incorporates arbitrary scalar covariates $\{\mathcal{X}_a, a = 1, \ldots, p_s\}$, functional covariates $\{X_a(v_a), a = 1, \ldots, p_f\}$, scalar-by-function interactions, and multiple levels of random effect covariates $\{Z_l^h, h = 1, \ldots, H; l = 1, \ldots, L_h\}$. In principle, our approach can also accommodate function-by-function interactions, but we omit that here. The general model can be written

$$\begin{aligned} y_i(t) = &\sum_{a=1}^{p_s} \mathcal{X}_{ia} B_a(t) + \sum_{a=1}^{p_f} \int_{v_a \in \mathcal{V}_a} X_{ia}(v_a) \boldsymbol{\beta}_a(v_a, t) dv_a \\ &+ \sum_{a_s=1}^{p_{s_I}} \sum_{a_f=1}^{p_{f_I}} \int_{v_{a_f} \in \mathcal{V}_{a_f}} \mathcal{X}_{ia_s} X_{ia_f}(v_{a_f}) \boldsymbol{\beta}_{a_s a_f}(v_{a_f}, t) dv_{a_f} + \sum_{h=1}^{H} \sum_{l=1}^{L_h} Z_{il}^h U_l^h(t) + E_i(t), \end{aligned} \tag{1.6}$$

where $B_a(t)$ are functional coefficients for scalar predictors, $\boldsymbol{\beta}_a(v_a, t)$ are function-on-function coefficient surfaces for functional predictors, $\boldsymbol{\beta}_{a_s a_f}(v_{a_f}, t)$ coefficient surfaces for the interaction of scalar covariate a_s and functional predictor a_f, and the random effects $U_l^h(t) \sim \mathcal{GP}(0, \Sigma_U^h)$. The multiple levels of random effects allow the model to handle various types of multi-level models needed to accommodate many complex designs commonly encountered in practice. Our code is capable of fitting this general model, although increasing number of functional predictors adds to the computational intensiveness of the sampler.

For the ERP data considered in Section 1.5, we include a discrete factor *image type* both as

a main effect as well as effect modifier for the functional predictor, which allows different functional intercepts and function-on-function regression surfaces for each image type, allowing us to investigate whether the brain responds differently to cigarette-related images and neutral, non-emotional images. See Model (1.10) in Section 1.5 for specification. Inference can then be performed on any number of desired statistics resulting from the model.

1.3 Posterior Functional Inference

Previous work in the function-on-function setting has focused solely on estimation or inference based on the construction of point-wise confidence intervals over the surface considering intervals that don't contain zero as significant (Scheipl, Staicu, and Greven, 2014). However, such an approach does not account for the inherent multiple testing problem from testing multiple locations within the coefficient surface. When applied to Bayesian credible intervals, we refer to this as the point-wise credible interval (PWCI) procedure. This unadjusted approach may lead to coefficients spuriously designated as significant. Thus we propose two posterior functional inference procedures aimed at flagging significant regions of a surface while controlling overall α, either using false discovery rate or experiment wise error rate, plus a Bayesian global test for testing whether the regression surface is identically zero..

First, we extend the Bayesian False Discovery Rate (BFDR) implemented by Morris et al. (2008) and Malloy et al. (2010) to the function-on-function setting. The BFDR is reliant upon the selection of δ-fold intensity change. Ideally this value is biologically motivated, however such a value may not necessarily exist or may be difficult to determine. Therefore, we also consider joint credible bands similar to those considered by Ruppert, Wand, and Carroll (2003) and introduce Simultaneous Band Scores (SimBaS), which are the minimum α required such that zero is excluded from the interval.

Suppose we have M MCMC samples. Let $\beta^{(m)}(v,t)$ be one realization of the posterior of the estimated surface for sample $m,\ m = 1,\ldots,M$. Then for a specific $v,\ v = 1,\ldots,V$, and $t,\ t = 1,\ldots,T$, we can consider the probability

$$P_{BFDR}(v,t) = Pr\left\{|\beta(v,t)| > \delta|y\right\} \approx \frac{1}{M}\sum_{m=1}^{M} 1\left\{\left|\beta^{(m)}(v,t)\right| > \delta\right\},$$

where δ is the pre-determined intensity change in the effect. To correct for the discrete nature of the MCMC we replace any $P_{BFDR}(v,t) = 1$ with the quantity $1-(2M)^{-1}$. The local FDR estimate for location (v,t) is then given by $1 - P_{BFDR}(v,t)$.

For a pre-specified global FDR-bound α, we flag the set of points (locations) satisfying $\psi = \{(v,t) : P_{BFDR}(v,t) \geq \nu_\alpha\}$. To obtain ν_α, we sort $\{P_{BFDR}(v,t), v = 1,\ldots,V, t = 1,\ldots,T\}$ in descending order across all sets of locations. This gives us the set $\left\{P_{(r)}, r = 1,\ldots,R\right\}$, where $R = V \times T$ or the ordered set of probabilities calculated above. We then define

$$\lambda = \max\left[r^* : \frac{1}{r^*}\sum_{r=1}^{r^*}\left\{1 - P_{(r)}\right\} \leq \alpha\right].$$

The cutoff for flagging significant coefficients is then $\nu_\alpha = P_{(\lambda)}$.

Alternatively and in the spirit of Ruppert, Wand, and Carroll (2003), consider constructing joint credible bands. A $100(1-\alpha)\%$ credible band of $\beta(v,t)$ must satisfy

$$Pr\left\{L(v,t) \leq \beta(v,t) \leq U(v,t)\ \forall\ v \in \mathcal{V}, t \in \mathcal{T}\right\} \geq 1-\alpha \tag{1.7}$$

where $L(v,t)$ and $U(v,t)$ are the lower and upper bounds respectively. It follows from Ruppert, Wand, and Carroll (2003) that an interval satisfying (1.7) is

$$I_\alpha(v,t) = \hat{\beta}(v,t) \pm q_{(1-\alpha)}\left[\widehat{\text{St.Dev}}\left\{\hat{\beta}(v,t)\right\}\right],$$

where $\hat{\beta}(v,t)$ and $\widehat{\text{St.Dev}}\left\{\hat{\beta}(v,t)\right\}$ are the mean and standard deviation for a given (v,t) taken over all M MCMC samples. The variable $q_{(1-\alpha)}$ is the $(1-\alpha)$ quantile taken over M of the quantity

$$Z^{(m)} = \underset{v \in \mathcal{V}, t \in \mathcal{T}}{\max}\left|\frac{\beta^{(m)}(v,t) - \hat{\beta}(v,t)}{\widehat{\text{St.Dev}}\left\{\hat{\beta}(v,t)\right\}}\right|.$$

These joint bands benefit from controlling for multiple testing in a strong experiment-wise fashion while also not requiring a pre-specified δ-fold intensity change as in the BFDR.

Now consider constructing $I_\alpha(v,t)$ for multiple levels of α and determining for each (v,t) the minimum α at which each interval excludes zero, denoted $P_{SimBaS}(v,t) = \min\{\alpha : 0 \notin I_\alpha(v,t)\}$, which can be directly computed by

$$P_{SimBaS}(v,t) = \frac{1}{M}\sum_{m=1}^{M} 1\left\{\left|\frac{\hat{\beta}(v,t)}{\widehat{\text{St.Dev}}\left\{\hat{\beta}(v,t)\right\}}\right| \geq Z^{(m)}\right\}. \tag{1.8}$$

We call these probabilities Simultaneous Band Scores or SimBaS. Similar to the BFDR and PWCI, we can select a specific α and flag (v,t) for which $P_{SimBaS}(v,t) < \alpha$ as significant, which is equivalent to checking if the joint credible intervals cover zero at a specific α-level. We can also compute global Bayesian p-values (GBPV), P_{GBPV} =min$_{v,t}\{P_{SimBaS}(v,t)\}$, a measure for testing the global null hypothesis that $\beta(v,t) = 0 \; \forall \; v \in \mathcal{V}, t \in \mathcal{T}$, when desired.

The BFDR, SimBaS, and GBPV can be computed for individual surfaces $\beta(v,t)$ or any transformation or contrast defined across surfaces. For example, in the two surface setting, interest focuses on applying the procedure to both $\beta_g(v,t),\ g = 0,1$, as well as the difference surface, $D(v,t) = \beta_1(v,t) - \beta_0(v,t)$. This allows us to detect differences between the two surfaces and flag where those differences occur.

1.4 Simulation

We generate data in two phases. First, we draw $\mathbf{x}_{ic}$, $\mathbf{u}_i$, and $\mathbf{e}_{ic}$. Second, we generate $\mathbf{y}_{ic}$ using $\mathbf{y}_{ic} = \mathbf{x}_{ic}\boldsymbol{\beta} + \mathbf{u}_i + \mathbf{e}_{ic}$, where $\boldsymbol{\beta}$ is one of four true surfaces of association. To generate predictor curves, random effects, and model errors, we use Gaussian Processes with autoregressive 1 [AR(1)] covariance structures. Estimates for parameters of the covariance of $\mathbf{x}_{ic}$ come from estimating autoregressive parameters from the output of one electrode from our ERP data. We assign pairs of curves to each subject to induce repeated measures and

consider three different sample sizes: $n = 25$, 50, and 100. Repeated measures brings the total number of observations up to $N = 50$, 100, and 200 respectively. We select parameters for the covariances of $\mathbf{u}_i$ and $\mathbf{e}_{ic}$ as $\sigma_E^2 = 0.1$ and $\rho_E = 0.5$ and $\sigma_U^2 = 0.05$ and $\rho_U = 0.75$ respectively. Prior to constructing $\mathbf{y}_{ic}$, we center and scale $\mathbf{x}_{ic}$ across i, c by time point so that the variance at each time point is 1.

We select true surfaces to mimic biologically plausible time varying associations. The top row of Figure 1.1 contains the heat maps of each surface. Each surface represents a different type of association, equations for which can be found in the Appendix. The ridge surface represents a relationship where the strongest association between $x(v)$ and $y(t)$ occurs along the line $v = t$. In other words, changes in $y(t)$ are associated with concurrent changes in $x(v)$. The lagged surface suggests a relationship where changes in $x(v)$ at a given time are associated with later changes in $y(t)$, but the strongest effect is delayed. The relationship between $x(v)$ and $y(t)$ in the immediate surface is similar to that in the lagged, however the strongest effect occurs immediately before dying off. Finally, the peak scenario demonstrates a setting where changes in $y(t)$ at a given time are associated with later changes in $x(v)$ and the association is characterized by a single peak.

For each surface, we generate 200 simulated data sets and draw posterior samples using a burn-in of 1000 followed by a chain of 1000 samples. We use Daubechies wavelets with four vanishing moments and three levels of decomposition. In preliminary simulations, zero-padding reduced edge effects better than symmetric-half point padding. Thus we implement zero-padding for all models. Motivated by the ERP data structure, we set the total number of time points in both time domains to be 225. For the wPC decomposition, we keep components accounting for 99.0% of the variability in $\mathbf{X}^W$. Averaged posterior estimates for each surface are found in the bottom row of Figure 1.1. Results from all three sample sizes were similar, thus we only present simulations for $n = 25, N = 50$ here. Results for $n = 100, N = 200$ can be found in the Appendix. For each dataset we also calculate root Mean Square Error (rMSE).

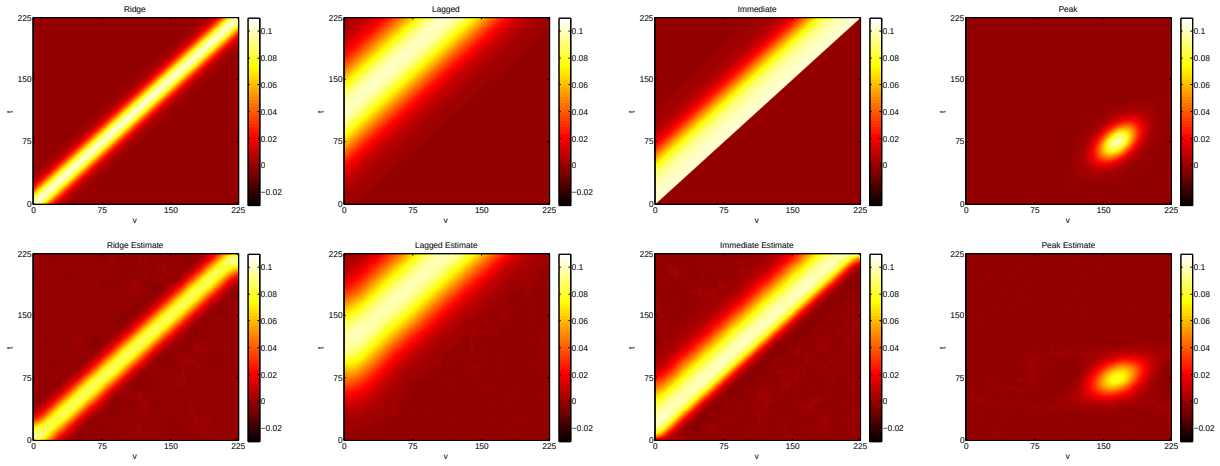


Figure 1.1: Heat maps of the true surfaces for simulation study are in the top row. The bottom row contains estimated surfaces for each simulated scenario based on a sample size of $n = 25$ with two measure per subject, $C_i = 2 \; \forall \; i$, for a total of $N = 50$ observations. Each surface is the average of the posterior estimate for the true surface based on 200 simulated datasets.

We also examine the performance of the BFDR, SimBaS, and GBPV procedures in simulation using a global α of 0.05. For the BFDR, we use a δ-intensity change of 0.05 which is roughly half the max signal from each surface. For comparison, we also generate unadjusted PWCIs. To evaluate the three procedures, we calculate false discovery rate, sensitivity, experiment-wise error rate (EWER), and type I error. Define false discovery rate, FDR_ϵ, as the number of flagged locations (v, t) with true value $\leq \epsilon$ divided by the total number of flagged locations. Next define the sensitivity, SEN_Υ, as the number of flagged locations (v, t) with true magnitude $> \Upsilon$ divided by the total number of locations with true magnitude $> \Upsilon$. EWER_ϵ is calculated as the proportion of simulated datasets with at least one falsely discovered location, i.e. a flagged location with true value $\leq \epsilon$. Type I error is calculated using a null simulation with true surface $\beta(v, t) = 0 \; \forall \; v \in \mathcal{V}, t \in \mathcal{T}$ and determining the proportion of simulated datasets with at least one location flagged as significant.

Figure 1.1 allows for direct comparison of each estimated surface to the truth. For all surfaces, we see the model performed quite well, effectively reconstructing all the true surfaces. Estimation improves as sample size increases. Not surprisingly, rMSE decreases

as sample size increases though even the smallest sample size produced small rMSEs. Heat maps containing the averaged set of flagged coefficients, ψ, for the BFDR and the average SimBa scores across datasets can be found in the Appendix. Both procedures correctly identified regions of elevated association in all four surfaces.

Table 1.1 displays both the average false discovery rate, FDR_ϵ, and the average sensitivity, SEN_Υ, for each scenario using $\epsilon = 0.01, 0.05$ and $\Upsilon = 0.05, 0.075$. For each procedure, we use $\alpha = 0.05$ to select the set of flagged coefficients. We can see that the BFDR and SimBaS procedures performs similarly well by both measures, though BFDR does better for a higher ϵ and Υ. While the PWCI has very good sensitivity, it comes at the cost of an inflated false discovery rate. EWER_ϵ is calculated using $\epsilon = 0.01$. Additionally, SimBaS controls experiment-wise type I error quite well at 0.05. While BFDR has a slightly low type I error of 0.04, PWCI has a very high value of 0.645. To assess P_{GBPV} we determine the percent of datasets under each scenario with $P_{GBPV} < 0.05$. In each scenario, all datasets have $P_{GBPV} < 0.05$.

These simulation results suggest our method performs well both in estimation and in inference. Even at the smallest sample size we considered, for this signal to noise ratio the model effectively reproduces the true surface. Both the BFDR and SimBaS capture the strongest regions of association without spuriously flagging too many non-significant coefficients. They also control well for type I error. Further, BFDR and SimBaS outperform the PWCI while maintaining reasonable sensitivity. Increasing sample size improves these facets of the model. Additional results, not included here nor in the Appendix, are available upon request.

Table 1.1: FDR, sensitivity, experiment-wise error rate (EWER), and type I error values by inference procedure. The BFDR use a δ intensity change of 0.05. To determine assessment values for SimBaS, a cutoff of $\alpha = 0.05$ was used. Likewise, the PWCI used 95% point-wise credible intervals to determine significant locations.

Measure	Surface	BFDR	SimBaS	PWCI
$FDR_{0.01}$	Lagged	0.06%	0.08%	5.80%
	Peak	0.48%	0.75%	22.9%
	Ridge	0.12%	0.19%	20.5%
	Immediate	2.25%	2.80%	20.9%
$FDR_{0.05}$	Lagged	5.74%	13.9%	44.7%
	Peak	4.01%	20.4%	73.5%
	Ridge	9.75%	15.6%	53.3%
	Immediate	5.74%	7.58%	38.1%
$SEN_{0.05}$	Lagged	98.1%	96.2%	99.9%
	Peak	64.9%	73.4%	99.9%
	Ridge	96.8%	93.4%	99.9%
	Immediate	97.9%	93.8%	99.9%
$SEN_{0.075}$	Lagged	99.9%	99.3%	100%
	Peak	94.4%	88.2%	99.9%
	Ridge	99.8%	97.6%	99.9%
	Immediate	99.9%	96.2%	99.9%
$EWER_{0.01}$	Lagged	7.00%	16.5%	100%
	Peak	4.50%	10.5%	100%
	Ridge	9.50%	49.0%	100%
	Immediate	100%	100%	100%
Type I Error	Null	4.00%	5.00%	64.5%

1.5 Application

1.5.1 Description of ERP Data Set

To demonstrate the features of the proposed model, we analyze data from the Department of Behavioral Sciences at the University of Texas M. D. Anderson Cancer Center. As part of a smoking cessation trial, researchers obtained Event Related Potentials (ERPs) at baseline for subjects viewing a series of images of different types, including neutral, emotional (positive and negative), and cigarette-related.

EEG was continuously recorded during image presentation and collected using a 129-channel Geodesic Sensor Net and amplified with AC-coupled high-input impedance (200 MΩ) amplifier (Geodesic EEG System 250; Electrical Geodesics, Inc., Eugene, OR) referenced to the Cz electrode. The time series were preprocessed as described in Versace et al. (2010a), with 0.1Hz high pass and 100Hz low pass filters, blink-corrected using spatial filtering, transformed to average reference, segmented into 900ms segments from 100ms before each image shown to 800ms after, obvious artifacts removed, and ERPs averaged across images for each image type per subject/electrode. After this processing, for each subject, we are left with functions of length 225 for each image type for all 129 electrodes.

Example curves recorded from 180 participants at electrode Cz (#129, in the middle of the crown of the head) during presentation of cigarette-related and neutral images can be seen in Figure 1.2 with the average over curves included in red. Curves under the other image-types are similar in appearance. The irregularity and localized spikiness of the raw curves motivates our use of wavelets in our modeling approach (Figure 1.2).

While many analyses are of interest for these data, in this paper we aim to characterize the time-varying relationship between ERP output from pairs of electrodes, focusing on two pairs in particular. The first pair is 55 and 129. Electrode 129, as previously mentioned, is positioned at the top of the head and electrode 55 is located directly behind it. We expect these two adjacent electrodes to be positively associated along the diagonal, $t = v$,

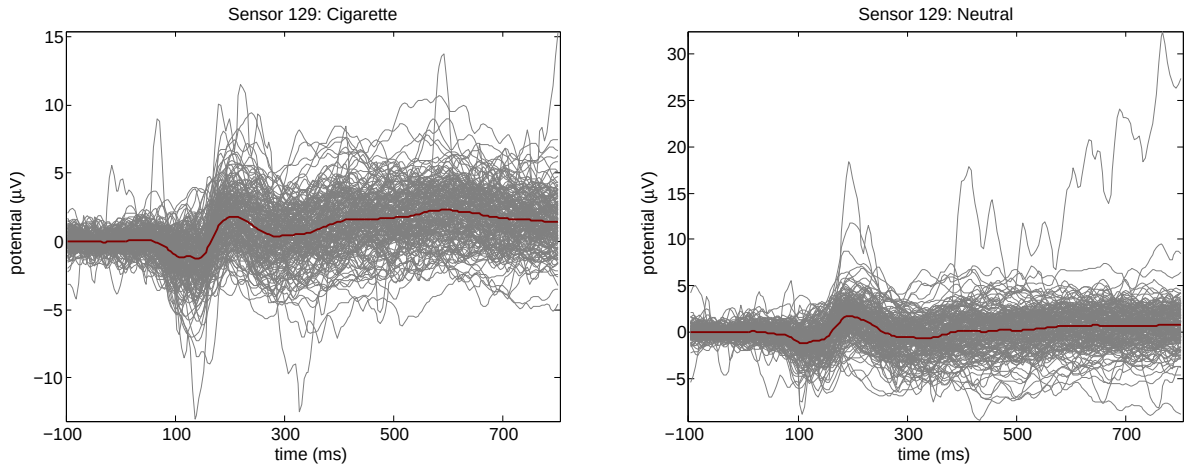


Figure 1.2: On the left, raw profile curves are plotted in gray with the mean in red from electrode 129 under the cigarette image condition. On the right, are raw curves and the mean from electrode 129 under the neutral image condition.

axis. The second pair is 75 and 11. Electrode 75 is an occipital electrode located at the back of the head while electrode 11 is at the front. Output from these two electrodes is expected to exhibit a negative correlation and thus we anticipate a negative association along the diagonal axis. For each pair of electrodes, we jointly model the association between the electrodes under both the neutral and cigarette image conditions resulting in a multilevel data structure. Thus for each model, subjects have four curves resulting from measurements from two electrodes while viewing two different image types.

1.5.2 Analysis

We fit two models to the data. In general, the model is given by

$$y_{icg}(t) = 1(g=0)\left[\alpha_0(t) + \int_{v\in\mathcal{V}} x_{ic0}(v)\beta_0(v,t)dv\right] \tag{1.9}$$

$$+ 1(g=1)\left[\alpha_1(t) + \int_{v\in\mathcal{V}} x_{ic1}(v)\beta_1(v,t)dv\right] + U_i(t) + E_{ic}(t), \tag{1.10}$$

where g denotes group membership, 0 for neutral, 1 for cigarette. For the first model we used Electrode 129 as the outcome function and Electrode 55 as the predictor function and for the second we used Electrode 11 as the outcome and Electrode 75 as the predictor.

In both models, inference was drawn on both image-specific surfaces, β_0 and β_1, as well as the difference surface $D(v,t) = \beta_1(v,t) - \beta_0(v,t)$. As in the simulation study, we used Daubechies wavelets with four vanishing moments and three levels of decomposition along with zero-padding. Prior to decomposition, we standardized both outcome and predictor functions by time. After DWT, the dimensions of the transformed functional outcomes from Electrodes 129 and 11 were both 360×245. After wPC, the dimensions of the transformed functional predictors were 360×72 for Electrode 55 and 360×62 for Electrode 75. We obtained 1000 posterior samples from the MCMC after a burn in of 1000. Spot checks of the trace plots of key parameters suggested MCMC convergence.

We considered inference for both models using all three procedures. For the BFDR procedure, we selected a global α of 0.05 when implementing BFDR on the difference surfaces. We choose a somewhat strict intensity change of $\delta = 0.05$ to focus on large differences between the surfaces. We also implemented BFDR on the image-specific surfaces in both models. There the α-level was reduced to 0.025 for each surface, however the intensity change, δ, remained at 0.05 so to only flag relatively large associations. Using the same intensity change for both models allows us to compare the two. For the PWCI, we also used $\alpha = 0.05$.

Figure 1.3 contains posterior means of all three surfaces for both models. Examination of the posterior estimates of the difference surfaces found in the first column of Figure 1.3 suggest little to no systematic difference between image type in both models. When we look at the image-specific surfaces in the model using electrodes 129 and 55 (top row, second and third column, Figure 1.3), we see an elevated ridge of association along the $t = v$ diagonal, which is the relationship we anticipated between these two adjacent electrodes. Note that this relationship is strongest in the first 300 ms in the ERP or 200 ms post picture presentation (image presentation occurred at $t = v = 0$), which corresponds to the initial response to viewing the image. Transitioning to the image-specific surfaces of the model using electrodes 11 and 75 (bottom row, second and third column, Figure 1.3), we see a valley of negative association along the $t = v$ diagonal that also begins to die out around

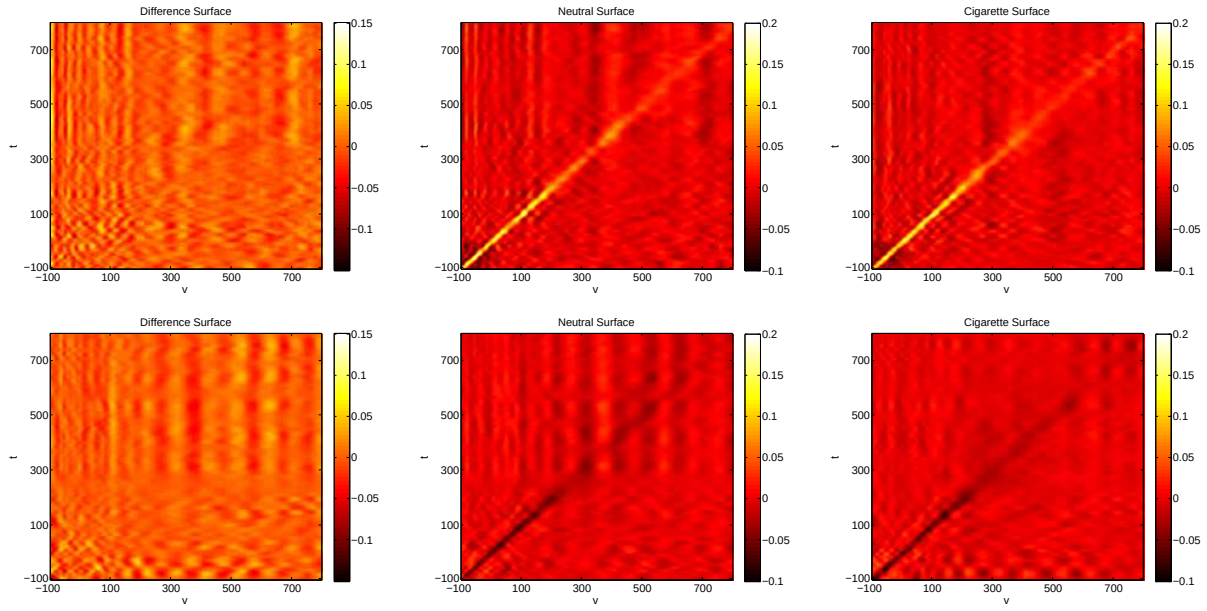


Figure 1.3: The top row contains surface estimates for the association between electrodes 129 and 55. Posterior surfaces comparing electrodes 11 to 75 are in the second row. The estimated posterior surface of the difference between cigarette and neutral is found in the first column. Group specific surface estimates are in the second and third columns, Neutral and Cigarette respectively. ERP output from electrode 129 is the response and the output from electrode 55 is the predictor for the first model and electrode 75 is the predictor of electrode 11 in the second model.

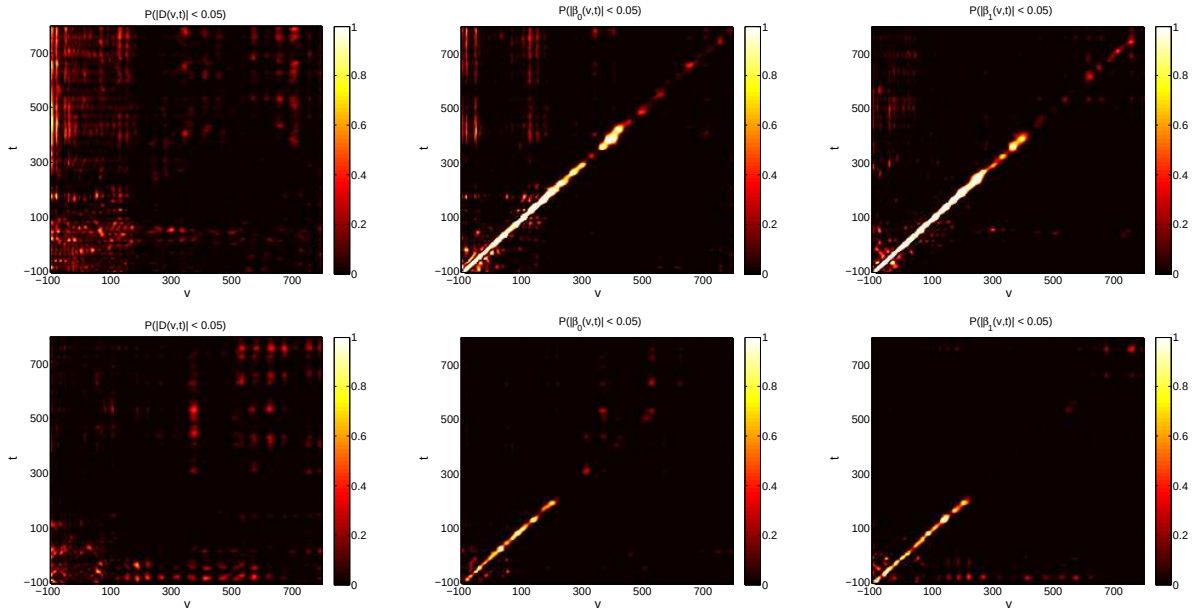


Figure 1.4: Heat maps containing the posterior probabilities from the BFDR procedure usinga δ intensity change of 0.05. Coefficients in white have a high probability of being greater than δ and thus likely to be included in ψ, the set of coefficients flagged as significant. Black coefficients have a low probability of being greater than δ and are thus less likely to be flagged as significant. The top row contains results from the model using electrodes 129 and 55 while the second row contains results from the model using electrodes 75 and 11.

200 ms to 300 ms past presentation. Once again, this is consistent with the expected relationship between these two electrodes.

Figure 1.4 contains results from the BFDR procedure on the difference surface for both models. Each heat map plots the posterior probabilities P_{BFDR}. We see that for both models, most locations have a low probability of being greater than δ. In fact, plotting ψ, we see no significantly flagged regions (see the Appendix), suggesting there is little evidence that the correlation across the two electrodes differs across image types. The second and third columns of Figure 1.4 show the application of the BFDR to the image-specific surfaces in each model, and again the heat maps plot the posterior probabilities P_{BFDR}. We see that the probabilities along the ridge are quite large suggesting that ridge of positive association is significant up until almost 300 ms past image presentation. However the negative association along the ridge we saw in the second model has lower probabilities

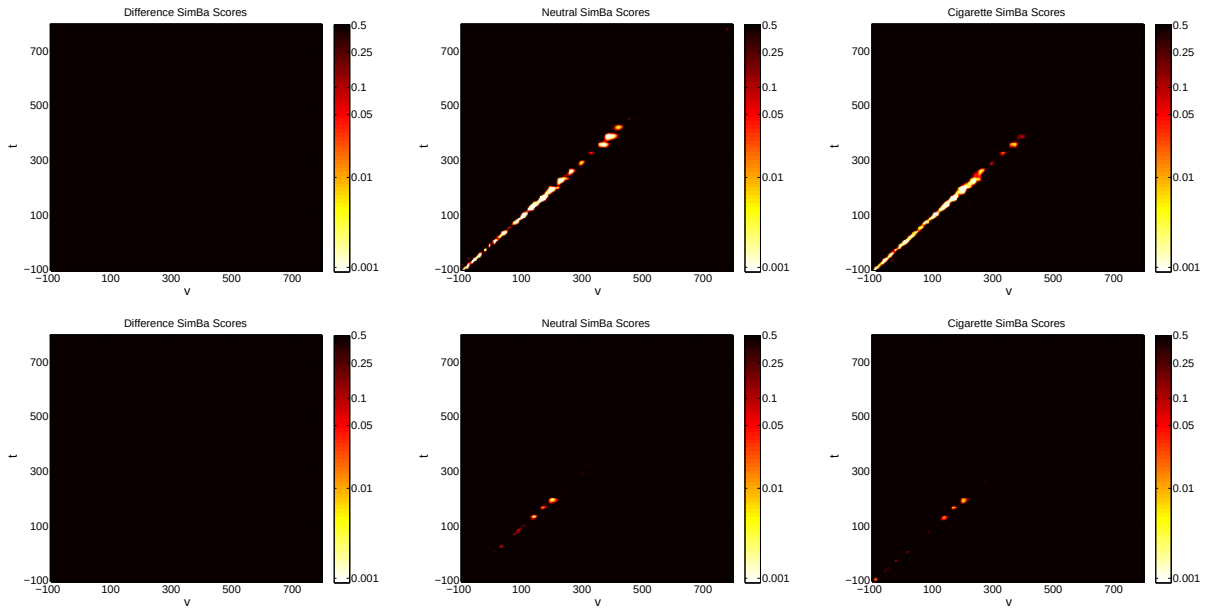


Figure 1.5: Heat maps containing the SimBa scores for each surface of both models. The top row contains results from the model using electrodes 129 and 55 while the second row contains results from the model using electrodes 75 and 11. Scores are plotted on the log-scale with the color axis on the exponential scale. White regions represent coefficients with low SimBa scores, black regions represent coefficients with high SimBa scores.

along the ridge at the $\delta = 0.05$ cut-off. Heat maps of the regions flagged as significant by the procedure are also found in the Appendix.

For the SimBaS procedure, we plot heat maps of the logged SimBa Scores in Figure 1.5. We see that for both difference surfaces, the SimBa scores are all relatively large (at least 0.5 or more), and the global Bayesian p-value for both is $P^D_{GBPV} = 0.5$, suggesting there is not enough evidence to conclude differences in the coefficient surfaces between image types. The second and third columns of Figure 1.5 show the SimBaS procedure applied to the image-specific surfaces. These heat maps are also plotted on the log-scale so to distinguish variations in small SimBa scores. For both models, we see evidence of a non-zero coefficient surface for each image type ($P^0_{GBPV} = P^1_{GBPV} = 0.001$ for the model using Electrodes 129 and 55, $P^0_{GBPV} = 0.001$ and $P^1_{GBPV} = 0.005$ for the model using Electrodes 11 and 75). Additionally, the SimBaS procedure detects the ridge of positive association in first model but only finds some of the negative associations in second.

Heat maps of flagged significant locations using PWCI can be found in the Appendix. Not surprisingly, the PWCI is more sensitive to minor variations in the surface where there appears to be no systematic association. While both BFDR and SimBaS found no significant locations in the difference surfaces, the PWCI flags a number of regions and also finds a number of significant locations in the image-specific surfaces that are off the $t = v$ axis while suggesting the association lingers longer. Given the results in the simulation studies, we interpret these results cautiously, as they may likely be spurious, and feel more confident in the multiplicity-adjusted inference from the BFDR and SimBaS procedures.

1.6 Discussion

Functional data analysis is an expanding field requiring more work to fill in gaps in the literature and build upon the general knowledge of the field. Previous work on function-on-function regression is limited. Here we present a general approach to function-on-function regression modeling which benefits from several attributes. First, our approach can use any basis function for $y(t)$ and $x(v)$ allowing us to handle functions of various types, including those with spiky and smooth features, and allowing us to parsimoniously model correlated residuals rather than assuming iid errors. Second, we get fully Bayesian inferences on all model quantities including point-wise credible intervals, posterior probabilities interpretable as Bayesian FDRs, joint credible intervals, and SimBaS that provide global and experiment-wise inferential quantities. Further generation of posterior predictive distributions is straightforward, so, for example, functional discriminant analysis can be performed (Zhu, Brown, and Morris, 2012). Third, our inference procedures correctly identify regions of elevated association without falsely flagging too many non-significant coefficients. Fourth, our method resides within the functional mixed model (FMM) framework as put-forth by Morris and Carroll (2006) that handles correlation between functions and random effects through random effect function distribution, and thus accounting for the various sources of variability in multi-level

models. Finally, the FMM framework also allows any combination of continuous and discrete scalar predictors, functional predictors, and their interactions, allowing function-on-function regression to be done in a much broader modeling context.

We demonstrated by simulation that our model performs well for realistic sample sizes and forms of functional association with fits improving as sample size increases. Simulations also show the BFDR and SimBaS procedures have better false discovery and type I error rates than the PWCI with comparable sensitivity. Our approaches for global inference and multiple-testing adjustment for Bayesian inference using BFDR, SimBaS, and GBPV are of general interest and can be used in other functional regression settings.

Our application displays the ability of the model to estimate the forms of the relationship of ERP output between different electrodes on the scalp. With the neighboring electrodes, a positive association was expected and seen along the diagonal axis $t = v$ while a negative association was expected and seen between electrodes on opposite sides of the scalp. Further, both our inference procedures were able to detect these associations as significant, even the one based on experiment wise error rate.

In summary, the function-on-function mixed model with basis-space modeling comprises a flexible approach to the function-on-function regression setting. The method performed well in both simulation and application. Further studies are needed to explore the model's performance in more complex settings, including non-functional components beyond a factor variable, incorporating multiple functional predictors, and various types of random effect correlation structures. Additionally, further examination of data reduction techniques could improve the modeling prowess of the method.

Acknowledgements

This work was supported by grants from the National Institutes of Health (ES007142, ES000002, ES016454, CA134294, CA160736, CA016672, DA032581).

References

Brumback, B. A., and Rice, J. A. (1998). Smoothing spline models for the analysis of nested and crossed samples of curves. *Journal of the American Statistical Association* **93**, 398–408.

Cardot, H., Ferraty, F., and Sarda, P. (1999). Functional linear model. *Statistics & Probability Letters* **45,** 11–22.

Chen, K. and Müller, H. G. (2012). Modeling Repeated Functional Observations. *Journal of the American Statistical Soceity* **107,** 1599–1609.

Cinciripini, P. M., Robinson, J. D., Karam-Hage, M., Minnix, J. A., Lam, C. Y., Versace, F., Brown, V. L., Engelmann, J. M., and Wetter, D. W. (2013). The effects of varenicline and bupropion-SR plus intensive counseling on prolonged abstinence from smoking, depression, negative affect, and other symptoms of nicotine withdrawal. *JAMA Psychiatry*, **70,** 522–533.

Gertheiss, J., Maity, A., and Staicu, A.-M. (2013). Variable Selection in Generalized Functional Linear Models. *Stat* **2,** 86–101.

Goldsmith, J., Bobb, J., Crainiceanu, C. M., Caffo, B., and Reich, D. (2011). Penalized Functional Regression. *Journal of Computational and Graphical Statistics* **20,** 830–851.

Goldsmith, J., Greven, S., and Crainiceanu, C. M. (2013). Corrected Confidence Bands for Functional Data Using Principal Components. *Biometrics* **69,** 41–51.

Harezlak, J., Coull, B. A., Laird, N. M., Magari, S. R., and Christiani, D. C. (2007). Penal-

ized solutions to functional regression problems. *Computational Statistics & Data Analysis* **51,** 4911–4925.

Hyvarinen, A., Karhunen, J., and Oja, E. (2001). *Independent Component Analysis*, Volume 26, Wiley-Interscience.

Ivanescu, A.E., Staicu, A.-M., Greven, S., Scheipl, F., and Crainiceanu, C. M. (2012). Penalized function-on-function regression. *Johns Hopkins University Working Paper*.

Johnstone, I. M., and Lu, A. Y. (2009). On Consistency and Sparsity for Principal Components Analysis in High Dimensions. *Journal of the American Statistical Society* **104,** 682–693.

Kim, K., Şentürk, D., and Li, R. (2011). Recent history functional linear models for sparse longitudinal data. *Journal of Statistical Planning and Inference* **141,** 1554–1566.

Malfait, N. and Ramsay, J. O. (2003). The historical functional linear model. *The Canadian Journal of Statistics* **31,** 115–128.

Malloy, E. J., Morris, J. S., Adar, S. D., Suh, H., Gold, D. R., and Coull, B. A. (2010). Wavelet-based functional linear mixed models: an application to measurement error-corrected distributed lag models. *Biostatistics* **11,** 432–452.

McLean, M. W., Hooker, G., Staicu, A.-M., Scheipl, F., and Ruppert, D. (2012). Functional Generalized Additive Models. *Journal of Computational and Graphical Statistics, to appear*.

Morris, J. S., Baladandayuthapani, V., Herrick, R. C., Sanna, P., and Gutstein, H. (2011). Automated analysis of quantitative image data using isomorphic functional mixed models, with application to proteomics data. *The Annals of Applied Statistics* **5,** 894–923.

Morris, J. S., Brown, P. J., Herrick, R. C., Baggerly, K. A., and Coombes, K. R. (2008). Bayesian Analysis of Mass Spectrometry Proteomic Data Using Wavelet-Based Functional Mixed Models. *Biometrics* **64,** 479–489.

Morris, J. S. and Carroll, R. J. (2006). Wavelet-based functional mixed models. *Journal of the Royal Statistical Society, Series B* **68,** 179–199.

Müller, H. G. and Yao, F. (2008). Functional Additive Models. *Journal of the American Statistical Society* **103,** 1543–1544.

Ramsay, J. O. and C. J. Dalzell (1991). Some tools for functional data analysis. *Journal of the Royal Statistical Society: Series B* **53,** 539–561.

Ramsay, J. O. and Silverman, B. W. (2005). *Functional Data Analysis* (Second ed.). Springer-Verlag.

Reiss, P. T., Huang, L., and Maarten, M. (2010). Fast Function-on-Scalar Regression with Penalized Basis Expansions. *The International Journal of Biostatistics* **6.**

Reiss, P. T. and Ogden, R. T. (2007). Functional principal components regression and functional partial least squares. *Journal of the American Statistical Association* **102,** 984–996.

Røislien, J., and Winje, B. (2012). Feature extraction across individual time series observations with spikes using wavelet principal component analysis. *Statistics in Medicine* **32,** 3660–3669.

Ruppert, D., Wand, M. P., and Carroll, R. J. (2003). *Semiparametric Regression*. Cambridge University Press.

Scheipl, F. and Greven, S. (2012). Identifiability in penalized function-on-function regression models. *University of Munich Technical Report* **125.**

Scheipl, F., Staicu, A.-M., and Greven, S. (2014). Functional Additive Mixed Models. *Journal of Computational and Graphical Statistics, to appear*. Available at *arXiv:1207.5947v5*.

Staicu, A.-M., Crainiceanu, C. M., Reich, D. S., and Ruppert, D. (2011). Modeling Functional Data with Spatially Heterogeneous Shape Characteristics. *Biometrics* **68,** 331-343.

Versace, F., Minnix, J. A., Robinson, J. D., Lam, C. Y., Brown, V. L., and Cinciripini, P. M. (2010a). Brain reactivity to emotional, neutral, and cigarette-related stimuli in smokers. *Addiction Biology* **16,** 296–307.

Versace, F., Robinson, J. D., Lam, C. Y., Minnix, J. A., Brown, V. L., Carter, B. L., Wetter, D. W., and Cinciripini, P. M. (2010b). Cigarette cues capture smokers attention: Evidence from event-related potentials. *Psychophysiology* **47,** 435–441.

Yao F., Müller H. G., and Wang J. L. (2005a). Functional Linear Regression Analysis for Longitudinal Data. *The Annals of Statistics* **33,** 2873–2903.

Zhu H., Brown, P. J., and Morris, J. S. (2011). Robust, adaptive functional regression in functional mixed model framework. *Journal of the American Statistical Association* **106,** 1167–1179.

Zhu H., Brown, P. J., and Morris, J. S. (2012). Robust classification of functional and quantitative image data using functional mixed models. *Biometrics* **68(4),** 1260–1268.

2. Bayesian Historical Functional Mixed Models for Repeated Measures

Mark J. Meyer[1], Elizabeth J. Malloy[2], Jeffrey S. Morris[3]
and Brent A. Coull[1]

[1]Department of Biostatistics, Harvard School of Public Health
[2]Department of Mathematics and Statistics, American University
[3]The University of Texas M.D. Anderson Cancer Center

2.1 Introduction

Recently, a number of articles treating the function-on-function regression setting have appeared in the literature (Scheipl and Greven, 2012; Ivanescu et al., 2012; Scheipl, Staicu, and Greven, 2014; Meyer et al., 2014). While these methods vary in approach, the core theme is to address the modeling situation where both the predictor and response are functions observed over a finely sampled grid. The regression surface in each model is, however, unconstrained. In other words, "outcome" effects occurring at time t can be associated both with "predictors" occurring prior to time t as well as after time t. Thus the unconstrained approach assumes no natural exposure-response relationship. When such a relationship is necessitated by the nature of the data, either the Historical Linear Functional Model (HFLM) as developed by Malfait and Ramsay (2003) and further explored by Harezlak et al. (2007) or the Recent History Functional Linear Model (RHFLM) proposed by Kim, Şentürk, and Li (2011) is more appropriate as both pre-supposes an exposure-response relationship. One such data setting comes from the environmental health literature where both macro and micro environment pollution levels are natural predictors of medically relevant biological processes.

Airborne Particulate Matter (PM) and Black Carbon (BC) are consistently shown in the literature to adversely affect cardiovascular health: Huang et al. (2012); Huttunen et al. (2012); Breysse et al. (2013); Neophytou et al. (2013) and references therein. Findings from these recent studies indicate that autonomic function is an important biological pathway and accumulating epidemiological evidence suggests that particles derived from micro-environment pollutants both occupational and non-occupational may be of particular concern. Two example studies in particular motivate our research. The Saint Louis Bus Study and the Boilermaker Study.

The Saint Louis Bus Study, described in Dubowsky et al. (2006), examined the relationship between both short-term and long-term pollution on autonomic function in the elderly. Subjects were repeatedly taken on field trips aboard a diesel powered bus with

continuous Holter electrocardiogram monitors to collect Heart Rate Variability (HRV) data, which is a measure of autonomic function. In this study design, investigators collected five minute averages of HRV on each subject from approximately 8:00AM on the morning of the trip to 7:00AM the following morning as well as five minute averages of traffic-related particles throughout the 48-hours surrounding each trip. Subjects were able to participate in up to four different field trips taken throughout the course of the study. Forty-four total subjects took part in the study. Anywhere from one to four curves were sampled on each subject resulting in 148 total profiles.

The Boilermaker Study, described first in Magari et al. (2001) and then in Cavallari et al. (2008), examined the effects of occupational exposure to airborne PM on boilermakers of varying occupation levels, apprentice and journeyman. Study subjects were fitted continuous Holter monitors to obtain HRV five minute averages as well as personal TSI Inc DustTrak device to measure personal exposure to $PM_{2.5}$ which is particulate matter less than 2.5μm in diameter. To measure HRV the authors used SDNN which is the Standard Deviation of N-to-N intervals over a five minute period. Study subjects were young and regularly exposed to residual oil fly ash (ROFA) that results as a byproduct of boiler construction. HRV and PM exposure curves from both the apprentice and journeyman boilermakers data are found in Figure 2.1.

Data from these two studies exhibit several features that, when combined, pose an interesting statistical problem. First, the pollution curves are natural predictors of HRV in that, for $v < t$ we expect a pollution measurement at time v might be associated with an HRV measurement at time t. However we do not expect the reverse association. Thus the function-on-function regression setting with an unconstrained surface is inappropriate, motivating the use of the HFLM. However direct application of the methods proposed in Malfait and Ramsay (2003), Harezlak et al. (2007), and Kim, Şentürk, and Li (2011) would result in over-smoothing as none of these methods are equipped to handle spiky and irregular functions. Additionally, the data is hierarchical in nature with 44 subjects taking a total of 148 trips which no previous approach can accommodate. To address

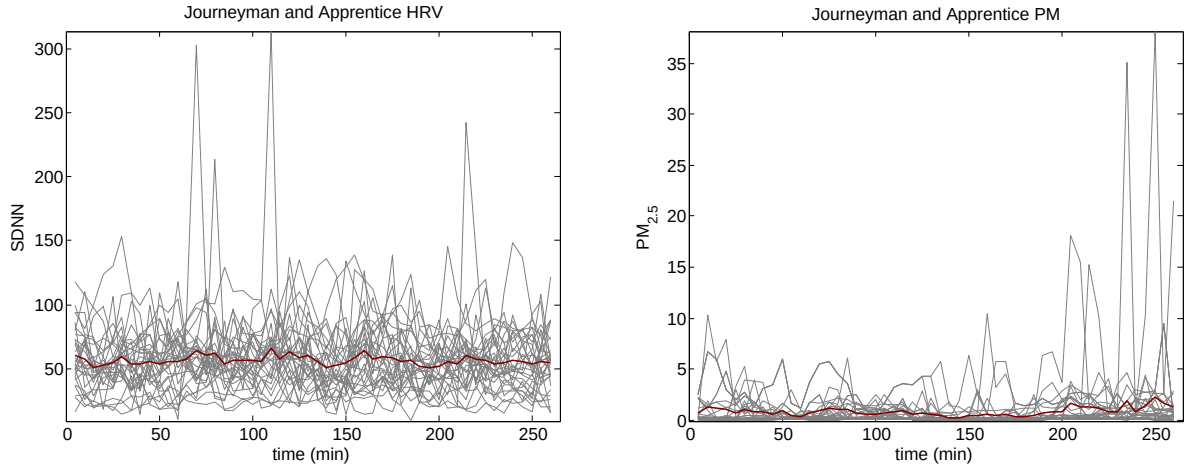


Figure 2.1: Boilermaker Study HRV and PM exposure profiles. Individual subject profiles are in gray, the mean of across time is in red. Time is measured as minutes from the start of measurement.

these statistical concerns as well as biological concerns arising from the study itself, such as whether freshly emitted particles are more toxic than older particles, we aim to develop methods that fully address time-varying effects of time-varying exposures on hierarchical functional outcomes. An important aspect of these methods is the development of inference procedures to summarize the health-exposure relationship after adjusting for potential confounders.

We thus propose a Bayesian Historical Functional Mixed Model which incorporates a novel use of the discrete wavelet-packet transformation (DWPT). We first motivate the use of DWPT over the discrete wavelet transformation (DWT) and demonstrate how the DWPT is uniquely configured to enforce the historical constraint. The modeling procedure will allow for both the historical functional component as well as potential scalar covariates of interest. Building off of the Functional Mixed Model framework of Morris and Carroll (2006), our method allows for the handling of correlations between functions and random effects. For inference, we propose the use of two posterior inferential procedures: the Bayesian False Discovery Rate (BFDR) as examined by Morris et al. (2008) and Malloy et al. (2010) and joint credible bands (Ruppert, Wand, and Carroll, 2003) alongside Simultaneous Band Scores as explored by Meyer et al. (2014).

What follows is the formulation of the Historical Functional Mixed Model in Section 2.2 which includes a discussion of the novel use of wavelet-packets for implementing the constraint. Section 2.3 presents the inference procedures. Sections 2.4 examines our method and inference procedures in simulation. Section 2.5 details the application of our approach to the Journeyman data from the Boilermaker Study. Finally Section 2.6 contains further discussion.

2.2 Historical Functional Mixed Models

We start by modifying the function-on-function mixed model proposed by Meyer et al. (2014) to incorporate the historical constraint. Thus for subject $i = 1, \ldots, n$, curve $c = 1, \ldots, C_i$, the historical functional linear mixed model is

$$y_{ic}(t) = \alpha(t) + \int_{\{v \leq t\}} x_{ic}(v)\beta(v,t)dv + U_i(t) + E_{ic}(t), \tag{2.1}$$

where $y_{ic}(t)$ and $x_{ic}(v)$ are predictor and response functions for subject i's cth curve with corresponding regression surface $\beta(v,t)$, $\alpha(t)$ is the intercept function, and $U_i(t)$ and $E_{ic}(t)$ are subject-specific and observation-specific Gaussian Process errors, $U_i(t) \sim \mathcal{GP}(0, \Sigma_U)$ and $E_{ic}(t) \sim \mathcal{GP}(0, \Sigma_E)$. This formulation also constitutes an extension of Malfait and Ramsay (2003) and Harezlak et al. (2007) to hierarchical functional data. Importantly, the restriction on the integral in Model (2.1) enforces the constraint that exposure recorded at a given time is uncorrelated with health outcomes collected earlier in time, given the full exposure time profile. That is, only exposures occurring before time t can affect health at time t.

Because data arrive sampled on a grid of discrete values, we use the discrete version of the model $\mathbf{y}_{ic} = \mathbf{x}_{ic}\boldsymbol{\beta} + \mathbf{u}_i + \mathbf{e}_{ic}$. In modeling, we recommend centering and scaling both the outcome and predictor functions. Thus, without loss of generality, we drop the intercept function from model formulation. Stacking the response vectors and predictor vectors

into matrices gives the model

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}U + \mathbf{E} \tag{2.2}$$

where for $N = \sum_i C_i$ total curves, $\mathbf{Y}$ is $N \times T$ and $\mathbf{X}$ is $N \times V$ since the response functions, $y_{ic}(t)$, are sampled at T equally spaced time points $t = [t_1, \cdots, t_T]'$, and the predictor functions, $x_{ic}(v)$, are sampled at V equally spaced time points $v = [v_1, \cdots, v_V]'$. Note the time domain of $x_{ic}(v)$ does not necessarily correspond to that of $y_{ic}(t)$. The constrained region of integration in Model (2.1) restricts the form of the functional regression coefficients so that $\beta(v_k, t_{k'}) = 0$ if $v_k > t_{k'}$. The discrete version of this requires that if $T = V$ and $t_1 = v_1, t_2 = v_2, \ldots, t_T = v_V$ then $\boldsymbol{\beta}$ is an upper triangular matrix of the form

$$\boldsymbol{\beta} = \begin{pmatrix} \beta(v_1, t_1) & \beta(v_1, t_2) & \cdots & \beta(v_1, t_m) \\ 0 & \beta(v_2, t_2) & \cdots & \beta(v_2, t_m) \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \beta(v_m, t_m) \end{pmatrix} \tag{2.3}$$

with zeros below the main diagonal. For the remainder of the model, $\mathbf{Z}$ is the $N \times n$ random intercept design matrix for the n study subjects. And $\mathbf{E}$ is an $N \times T$ matrix and assumed to come from a Gaussian Process. As we discuss further below, a primary goal of this research is to conduct estimation and inference while enforcing the upper triangular constraint in $\boldsymbol{\beta}$.

2.2.1 Model Formulation with Wavelets

Meyer et al. (2014) propose a generalized basis expansion for the function-on-function form of Model (2.2) where $\boldsymbol{\beta}$ is unconstrained. Indeed we could consider a variety of basis expansions for modeling provided that the historical constraint is maintained. Malfait and Ramsay (2003) consider triangular basis functions to enforce the lower triangle constraint. Likewise, Harezlak et al. (2007) examines penalization both corresponding to LASSO and spline methodology on similar triangular basis functions. Kim, Şentürk, and Li (2011) use B-spline basis functions and suggest the possible use of Fourier, truncated power, and Eigen basis functions. These approaches produce relatively smooth estimates

of the historical surface and thus may not be well suited to spiky and irregular functions. A natural choice of basis for such data is a wavelet basis.

First, apply a DWT separately to each row of $\mathbf{Y}$ and to each row of $\mathbf{X}$. Performing this transformation is equivalent to the post-multiplication of the approximately orthonormal projection matrices resulting from the DWT. Let W_y and W_x denote those matrices. Further let those transformation be indexed by scales $j = 1, \ldots, J^y$ and $s = 1, \ldots, S^x$ and locations $k = 1, \ldots, K_j^y$ and $\ell = 1, \ldots, L_s^x$ in the y and x wavelet-spaces respectively. Then for the DWT decompositions $\mathbf{Y} = \mathbf{Y}^W W_Y$, $\mathbf{X} = \mathbf{X}^W W_X$, $\boldsymbol{\beta} = W_X' \boldsymbol{\beta}^W W_Y$, $U = U^W W_Y$ and $\mathbf{E} = \mathbf{E}^W W_Y$ wavelet-space model is given by

$$\mathbf{Y}^W W_Y = \mathbf{X}^W W_X W_X' \boldsymbol{\beta}^W W_Y + \mathbf{Z} U^W W_Y + \mathbf{E}^W W_Y. \tag{2.4}$$

Noting that W_Y and W_X are orthogonal and post-multiplying by W_Y', Model (2.4) reduces to $\mathbf{Y}^W = \mathbf{X}^W \boldsymbol{\beta}^W + \mathbf{Z} U^W + \mathbf{E}^W$. Posterior estimate could then be sampled using an appropriate MCMC procedure. Noting that elements of $\boldsymbol{\beta}$ can be indexed by their corresponding locations and scales, priors in the wavelet-space might take the form of mixture prior such as

$$\beta^W_{(s\ell,jk)} \sim \ \gamma_{(s\ell,jk)} \mathcal{N}(0, \tau_j) + (1 - \gamma_{s\ell,jk}) d_0, \ \ \gamma_{(s\ell,jk)} \sim \mathcal{B}(\pi_j)$$

where d_0 is a point-mass distribution at zero and τ_j and π_j are regularization parameters (for further details see Meyer et al. (2014)). However this specification is for the unconstrained model. To enforce the historical constraint on $\boldsymbol{\beta}$ while using wavelets, we have two options to consider: the first is involves iterating between the wavelet-space and the data space, the second involves establishing a relationship between the time domain and the scales and locations in the wavelet domain.

To enforce the constraint, one approach could be to fit the fully functional model and at each step of iteration, project the estimates of $\boldsymbol{\beta}$ back into the data space setting $\beta(v, t) = 0$ for $v > t$. After enforcing the constraint, the estimates would then need to be projected back into wavelet-space to properly update the sampler. This approach suffers from several flaws. The first is computational intensity. Performing an additional inverse discrete

wavelet transformation (IDWT) and DWT at each step of the chain could dramatically impact computation time. The second, a more deleterious to model fitting, is that this approach requires the estimation of the whole surface which could result in a bleed-over effect whereby estimates are unduly influenced by the estimation of coefficients not satisfying the historical constraint. And finally, if the ultimate goal is estimation of the upper triangle only, the procedure wastes power on the estimation of coefficients below the triangle.

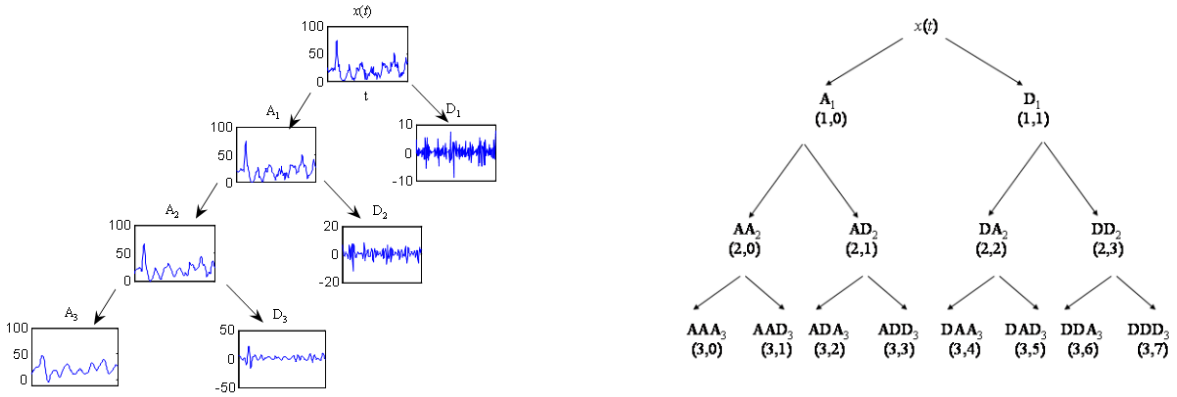
Another approach is to establish a relationship between the time domains in the data space and the scale and location coefficients in the wavelet-space. In this way, the constraint could be enforced in the wavelet-space and only the desired coefficients and parameters would be sampled. Thus we could select a prior on the wavelet coefficients $\beta^W_{(s\ell,jk)}$ that reflects the time-domain restriction in Model (2.3). Percival and Walden (2000) noted that the wavelet coefficients can be approximately associated with a specific time based on its location and scale, so that a coefficient with respect to scale j and location k is associated with $t_{jk} = (2k+1)2^{j-1} - \frac{1}{2}$. Likewise in the $s\ell$ dimension, $v_{s\ell} = (2\ell+1)2^{s-1} - \frac{1}{2}$. Thus, based on this approximate connection, one could consider the priors

$$\beta^W_{(s\ell,jk)} \sim \ 1(v_{s\ell} \leq t_{jk})\gamma_{(s\ell,jk)}\mathcal{N}(0,\tau_j) + (1-\gamma_{s\ell,jk})d_0, \ \ \gamma_{(s\ell,jk)} \sim \mathcal{B}(\pi_j)$$

where $1(v_{s\ell} \leq t_{jk})$ is 1 if its argument is true and 0 otherwise. However, this coefficient-time connection is only approximate. Because each wavelet coefficient is associated with a specific scale and location, each wavelet coefficient is actually associated with a specific *set* of times in dimension v and t which causes incorrect zeroing of the $\beta(v,t)$ surface. To solve this constraint problem, we propose a novel use of wavelet-packets.

2.2.2 Historical Constraint via Wavelet-Packets

Given space constraints, and to avoid excessive notation, we illustrate wavelet-packets in a heuristic fashion but note that full rigor can be found in Percival and Walden (2000) and Misiti et al. (2007). Consider a generic 1-dimensional function, $x(t)$. The DWT de-



(a) Discrete Wavelet Transformation (b) Discrete Wavelet-packet Transformation

Figure 2.2: (a) Decomposition of a function $x(t)$ into three levels using DWT, $x(t) = A_3 + D_3 + D_2 + D_1$. (b) Graphical representation of the decomposition of a function into three levels using DWPT, $x(t) = AAA_3 + AAD_3 + ADA_3 + ADD_3 + DAA_3 + DAD_3 + DDA_3 + DDD_3$.

composes $x(t)$ into an approximation and successive levels of detail Figure 2.2a. For instance, a 3-level decomposition of the row vector $x(t)$ using the DWT would start with a decomposition into an approximation A_1 and a detail component D_1. The second level of decomposition takes the approximation piece and further decomposes that into approximation and detail components, so that $x(t) = A_2 + D_2 + D_1$. For the third level of decomposition the approximation A_2 is split, giving $x(t) = A_3 + D_3 + D_2 + D_1$.

Wavelet packets are found in a similar manner as the DWT except at each stage *both* the approximation and the detail components are further decomposed, Figure 2.2b. The first stage of the DWPT looks the same as above for the DWT, as the function can be represented as $x(t) = A_1 + D_1$. For the second stage both the detail and approximation are decomposed, yielding: $x(t) = AA_2 + AD_2 + DA_2 + DD_2$. The third level of decomposition gives the final representation in Figure 2.2b. The wavelet coefficients therefore relate to a single level, yielding $x(t) = AAA_3 + AAD_3 + ADA_3 + ADD_3 + DAA_3 + DAD_3 + DDA_3 + DDD_3$. There are 2^L groupings of wavelet coefficients at the level L decomposition. Ordinarily, wavelet packets have been used to find an optimal decomposition of a function based on different detail/approximation combinations from the wavelet packet

tree (Misiti et al., 2007). For our purposes we are not interested in any optimal representation using the packets but in the final decomposition at a given level, L. Rather, when using the 2-dimensional DWPT for the regression surface $\boldsymbol{\beta}$, preservation of the constraint in Model (2.3) now follows directly because each s and j combination represents nodes at the same level of decomposition so that ℓ and k are associated with their corresponding time intervals v and t, respectively. Therefore $\beta(v,t) = 0$ for $v > t$ can be better approximated using the DWPT by setting $\beta^W_{(s\ell,jk)} = 0$ if $\ell > k$.

To illustrate proof of concept, consider the images in Figure 2.3 as an example of a hypothetical $\beta(v,t)$ function. It is a 256×256 pixel image defined for v and $t = 1, \cdots, 256$ where v runs along the horizontal axis and t the vertical axis. The top left figure of Figure 2.3 displays the original image which is true to the historical constraint, $\beta(v,t) = 0$ for $v > t$ where blue corresponds to $\beta(v,t) = 0$. The top right image in Figure 2.3 shows the results from the naïve approximate DWT restriction using a 3-level 2-dimensional DWT with the Haar wavelet family. After restricting the coefficients with the constraint $1(v_{s\ell} \leq t_{jk})$, the coefficients were transformed via 2-dimensional IDWT back into the time (v and t) domain. Here we see considerable distortion along the edge of the constraint. Further distortion can be seen in the upper triangle particularly as the coefficients decrease. A closer examination of diagonal shows an example of "ghosting" where coefficients that should be set to zero are not. Thus the constraint in the wavelet pace does not properly enforce the constraint.

Conversely, the bottom left figure in Figure 2.3 shows the constraint and reconstruction using wavelet packets, again using a 3-level decomposition with the Haar wavelet family. In the wavelet-packet space, we apply the given restriction, $\beta^W_{s\ell,jk} = 0$ if $\ell > k$. The result is an image that is essentially identical to the original image. Use of other wavelet families, such as Daubechies for two or more vanishing moments, results in greater distortion along the edge. This likely due in part to the padding inherent to other wavelet families. We suggest the use of the Haar wavelets as they maintain the edge better. Similar, a dyadic signal is required as padding can lead to distortions along the constraint. The

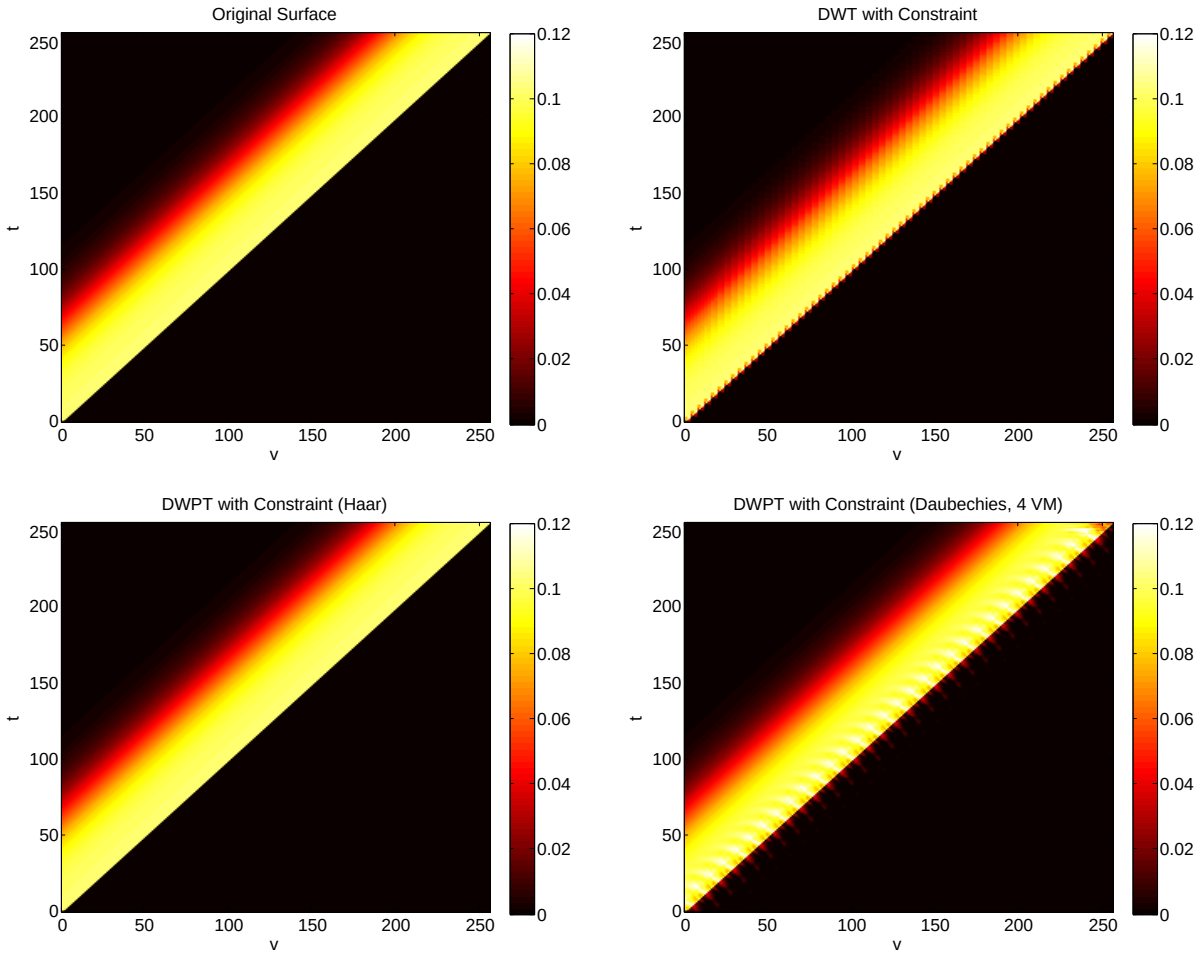


Figure 2.3: Proof of concept of the historical constraint. Top left: original image. Top right: decomposed and reconstructed original image with constraint in wavelet space. Bottom left: decomposed and reconstructed original image with constraint in wavelet-packet space using Haar wavelets. Bottom right: wavelet-packet space proof of concept using Daubechies wavelets with 4 vanishing moments.

point here is that the DWPT faithfully retains the features of the regression surface while enforcing the upper triangular constraint. But if we use Daubechies wavelets, as depicted in the bottom right corner of Figure 2.3, we see ghosting to either side of the constraint. Thus only the Haar wavelets maintain the constraint.

2.2.3 Model Formulation with Wavelet-packets and Thresholding

Now perform the DWPT to each row of $\mathbf{Y}$ and $\mathbf{X}$. The resulting decompositions have the form $\mathbf{Y} = \mathbf{Y}^{W_P} W_{P,Y}$ and $\mathbf{X} = \mathbf{X}^{W_P} W_{P,X}$ where $W_{P,Y}$ and $W_{P,X}$ are orthogonal matrices containing the wavelet packet basis functions. Then for the two dimensional decomposition on $\boldsymbol{\beta} = W'_{P,X} \boldsymbol{\beta}^{W_P} W_{P,Y}$ Model (2.2) in the wavelet-packet space is

$$\mathbf{Y}^{W_P} W_{P,Y} = \mathbf{X}^{W_P} W_{P,X} W'_{P,X} \boldsymbol{\beta}^{W_P} W_{P,Y} + \mathbf{Z} U^{W_P} W_{P,Y} + \mathbf{E}^{W_P} W_{P,Y}$$

for $U = U^{W_P} W_{P,Y}$ and $\mathbf{E} = \mathbf{E}^{W_P} W_{P,Y}$. Post-multiplying by $W_{P,Y}$ and recognizing the orthogonality of the wavelet-packet basis matrices, this model reduces to $\mathbf{Y}^{W_P} = \mathbf{X}^{W_P} \boldsymbol{\beta}^{W_P} + \mathbf{Z} U^{W_P} + \mathbf{E}^{W_P}$. While the decompositions have changed, the final form is very similar. Further this model is of the same as that formulated in Meyer et al. (2014). Indeed, we could proceed with their modeling procure from here using wavelet-packets instead of wavelets and wavelet-Principal Components (wPC). However, their approach does not enforce the desired historical constraint which we implement via prior specification.

The priors on the wavelet-packet model parameters are similar to the wavelet-space priors mentioned in Section 2.2.1 but with a slight modification. Instead of restricting based on $v_{s\ell}$ and t_{jk}, we restrict in wavelet-packet space if $\ell > k$. Thus our prior on the elements of $\boldsymbol{\beta}^{W_P} = \left[\beta^{W_P}_{(s\ell,jk)}\right]$ is

$$\beta^{W_P}_{(s\ell,jk)} \sim \; 1(\ell \leq k)\gamma_{(s\ell,jk)} \mathcal{N}(0, \tau_{jk}) + (1 - \gamma_{s\ell,jk}) d_0, \;\; \gamma_{(s\ell,jk)} \sim \mathcal{B}(\pi_{jk}),$$

the commonly used Gaussian-point mass mixture distribution. Thus we only sample coefficients from the space of interest.

Remaining priors are consistent with Meyer et al. (2014) where regularization parameters were assumed to come from an inverse-gamma distribution and a beta distribution for τ_{jk} and π_{jk} respectively. Hyper-parameters for both are fixed in both and based on the data. The Monte Carlo Markov Chain (MCMC) algorithm described in supplemental material to Meyer et al. (2014) need only be slightly modified to generate posterior samples of parameters. In particular, when the surface coefficients are sampled, we modify their procedure to maintain the historical constraint. In other words, we only sample wavelet-packet place coefficients satisfying the constraint $\ell \leq k$.

One computational issue discussed by Meyer et al. (2014) is the need for a data reduction in the X space. The algorithm becomes computationally intensive for large V. In their formulation, the authors used wPCs keeping columns containing 99.9% of the variability in $\mathbf{X}$. The wPC decomposition involves first performing a DWT on $\mathbf{X}$ and then performing a singular value decomposition on $\mathbf{X}^W$. A major benefit of this approach, not only in speeding up computation time, is the resulting denoising achieved. Indeed, removing columns for the wPC transformed $\mathbf{X}$ corresponds to thresholding the coefficients corresponding to those columns. The problem with implementing that basis in the historical framework is that the PCA breaks the relationship between packet location ℓ and time which is crucial for the implementation of the constraint. To remedy the computational concerns and simultaneously achieve large scale denoising, we propose a hard thresholding procedure.

Hard thresholding is a standard procedure in wavelet regression when DW transformed signal is scalar, i.e. non-functional. A variety of thresholds exist and detailed discussions of each can be found in Percival and Walden (2000) and Nason (2008). Hard thresholding involves picking a cut-off value and setting coefficients smaller than the cut-off to zero. In other words, given the DWT coefficients $\mathbf{d}^*$, a hard thresholding approach only keeps coefficients satisfying $\mathbf{d}^* > \lambda^H$. Donoho and Johnstone (1994) introduced the universal threshold for wavelet shrinkage which has the form $\lambda^u = \sigma\sqrt{2\log(n_d)}$ where n_d is the number of observations and σ is some estimate of the standard deviation of the noise. If $\mathbf{d}^*$ is variance 1 Gaussian noise than the threshold becomes just $\lambda^u = \sqrt{2\log(n_d)}$.

A straight forward implementation of the universal threshold is not possible in our context as we have what amounts to a sample coefficients from DWPTs performed on the rows of $\mathbf{X}$. Thus we propose a modification to the procedure where we threshold on the variance, which in the wavelet setting amounts to thresholding on the energy contributed. We first take the variance of the DWP transformed $\mathbf{X}$, $\mathbf{X}^{W_P}$. We then only keep columns whose variance is above the universal threshold. In practice, this amounts to removing close to seven-eighths of columns of $\mathbf{X}^{W_P}$ which achieves both the desired data reduction and denoting.

2.3 Posterior Functional Inference

Previous examinations of the HFLM did not discuss inferential procedures preferring instead to develop measures of model fit (Malfait and Ramsay, 2003; Harezlak et al., 2007). In the function-on-function literature, Meyer et al. (2014) propose the use of a Bayesian False Discovery Rate (BFDR) procedure also used by Morris et al. (2008) and Malloy et al. (2010) and joint credible bands as discussed in Ruppert, Wand, and Carroll (2003) alongside the calculation of Simultaneous Band Scores (SimBaS). For our historical functional mixed model we suggest the use of both.

The BFDR procedure begins by utilizing the MCMC samples to determine the posterior probability of a given coefficient being greater than a δ-fold intensity change. Once these values are determined they are ranked and a cut-off selected to control the overall FDR at a pre-specified global α-bound. Suppose we have M MCMC samples and $\beta^{(m)}(v,t)$ is one draw from the posterior estimated surface. Then for $\{v \in \mathcal{V} \text{ and } t \in \mathcal{T} \text{ s.t. } v \leq t\}$, we find the probability

$$P_{BFDR}(v,t) = Pr\left\{|\beta(v,t)| > \delta | y\right\} \approx \frac{1}{M}\sum_{m=1}^{M} 1\left\{\left|\beta^{(m)}(v,t)\right| > \delta\right\}.$$

In other words, we calculate P_{BFDR} over coefficients satisfying the constraint. Then we flag the set of locations on the historical surface satisfying $\psi = \{(v,t) : P_{BFDR}(v,t) \geq \phi_\alpha\}$

where ϕ_α is determined by first ranking the values of P_{BFDR} in descending order across all locations to obtain the set $\left\{P_{(r)}, r = 1, \ldots, R\right\}$ where R is the total number of coefficients satisfying the historical constraint. Then define $\lambda = \max\left[r^* : \frac{1}{r^*}\sum_{r=1}^{r^*}\left\{1 - P_{(r)}\right\} \leq \alpha\right]$. The cut-off for flagging coefficients as significant is then $\phi_\alpha = P_{(\lambda)}$. However, the BFDR relies upon a δ-intensity change. The choice δ may not be obvious in certain data situations. Thus we propose the use of the SimBaS procedure.

SimBaS begins by first constructing joint credible bands in the spirit of Ruppert, Wand, and Carroll (2003). A $100(1-\alpha)$% credible band of $\beta(v,t)$ must satisfy

$$Pr\left\{L(v,t) \leq \beta(v,t) \leq U(v,t)\ \forall\ v \in \mathcal{V}, t \in \mathcal{T} \text{ s.t. } v \leq t\right\} \geq 1 - \alpha \tag{2.5}$$

where $L(v,t)$ and $U(v,t)$ are the corresponding upper and lower band bounds. An interval satisfying Model (2.5) is given by $I_\alpha(v,t) = \hat{\beta}(v,t) \pm q_{(1-\alpha)}\left[\widehat{\text{St.Dev}}\left\{\hat{\beta}(v,t)\right\}\right]$ where $\hat{\beta}(v,t)$ and $\widehat{\text{St.Dev}}\left\{\hat{\beta}(v,t)\right\}$ are the posterior mean and standard deviation respectively and $q_{(1-\alpha)}$ is the $(1-\alpha)$ quantile taken over M of the quantity

$$\max_{v \leq t}\left|\frac{\beta^{(m)}(v,t) - \hat{\beta}(v,t)}{\widehat{\text{St.Dev}}\left\{\hat{\beta}(v,t)\right\}}\right|.$$

We construct $I_\alpha(v,t)$ for multiple α-levels and find the minimum α at which each interval excludes zero. These values, $\min\{\alpha : 0 \notin I_\alpha(v,t)\}$, we call the SimBa Scores and denote with P_{SimBaS}. We can consider SimBa scores to essentially be Bayesian p-values testing the null that $\beta(v,t) = 0$ for a specific location (v,t). Via their construction, the scores account for multiplicity of testing and benefit, in contrast to the BFDR, by not requiring specification of a δ-fold intensity change. To determine a set of flagged significant coefficients similar to ψ, we could select a specific α and flag all coefficients whose SimBaS fall below that bound. Meyer et al. (2014) demonstrates the properties of these procedures in the function-on-function regression setting and found both to be preferable to an unadjusted, point-wise interval approach. Additionally, they propose the Global Bayesian P-value (GBPV) as a test of $\beta(v,t) = 0\ \forall\ t \in \mathcal{T}, v \in \mathcal{V}$. The GBPV is calculated as the minimum SimBa Score across all (v,t) such that $v \leq t$.

2.4 Simulation

Here we present the details of a simulation study evaluating estimation both with and with out repeated measures. In particular, we generate data first based on the Saint Louis Bus Study to evaluate the methodology in the presence of unbalanced repeated measures. We next generate data without repeated measures but for a large sample size to demonstrate the large sample properties of the method. Data for the repeated measures scenarios was generated using

$$\mathbf{y}_{ic} = \mathbf{x}_{ic}\boldsymbol{\beta} + \mathbf{u}_i + \mathbf{e}_{ic}$$

where $\mathbf{x}_{ic}$ is generated for each subject from a multivariate normal distribution with a first oder auto-regressive, AR(1), covariance matrix. This data generation is consistent with the assessment of function-on-function regression seen in Meyer et al. (2014). For the repeated measures data generation, error terms were generated from mean zero multivariate normal distributions. An AR(1) covariance matrix was used for each with $\sigma^2_{\mathbf{u}} = 0.05$, $\rho_{\mathbf{u}} = 0.75$, $\sigma^2_{\mathbf{e}} = 0.1$, and $\rho_{\mathbf{e}} = 0.5$. The error terms for the large sample generation remained the same as those used in the repeated measures setting.

For both assessments, repeated measures and large sample, two hundred data sets were generated under each of three biologically relevant scenarios:

$$\frac{584}{10000}\frac{1-(256-t)}{256}\frac{1}{\sqrt{(2\pi)(0.01)}}\exp\left[-\frac{1}{(2)(0.01)}\left(\frac{t}{256}-\frac{v}{256}-0.5\right)^2\right] \text{ (cumulative)},$$

$$\frac{292}{10000}\frac{1}{\sqrt{(2\pi)(0.01)}}\exp\left[-\frac{1}{(2)(0.01)}\left(\frac{t}{256}-\frac{v}{256}-0.5\right)^2\right] \text{ (lagged)},$$

$$\frac{1029}{10000}\left[1-\frac{1}{1+\exp\left(\frac{0.25-\frac{t}{256}+\frac{v}{256}}{0.05}\right)}\right] \text{ (immediate)}.$$

Each true surface, displayed in the right column of Figure 2.4, had the constraint enforced to ensure that for $v > t$, the true values were set exactly to zero. The immediate effect depicts a scenario in which an exposure has an immediate impact on the outcome that

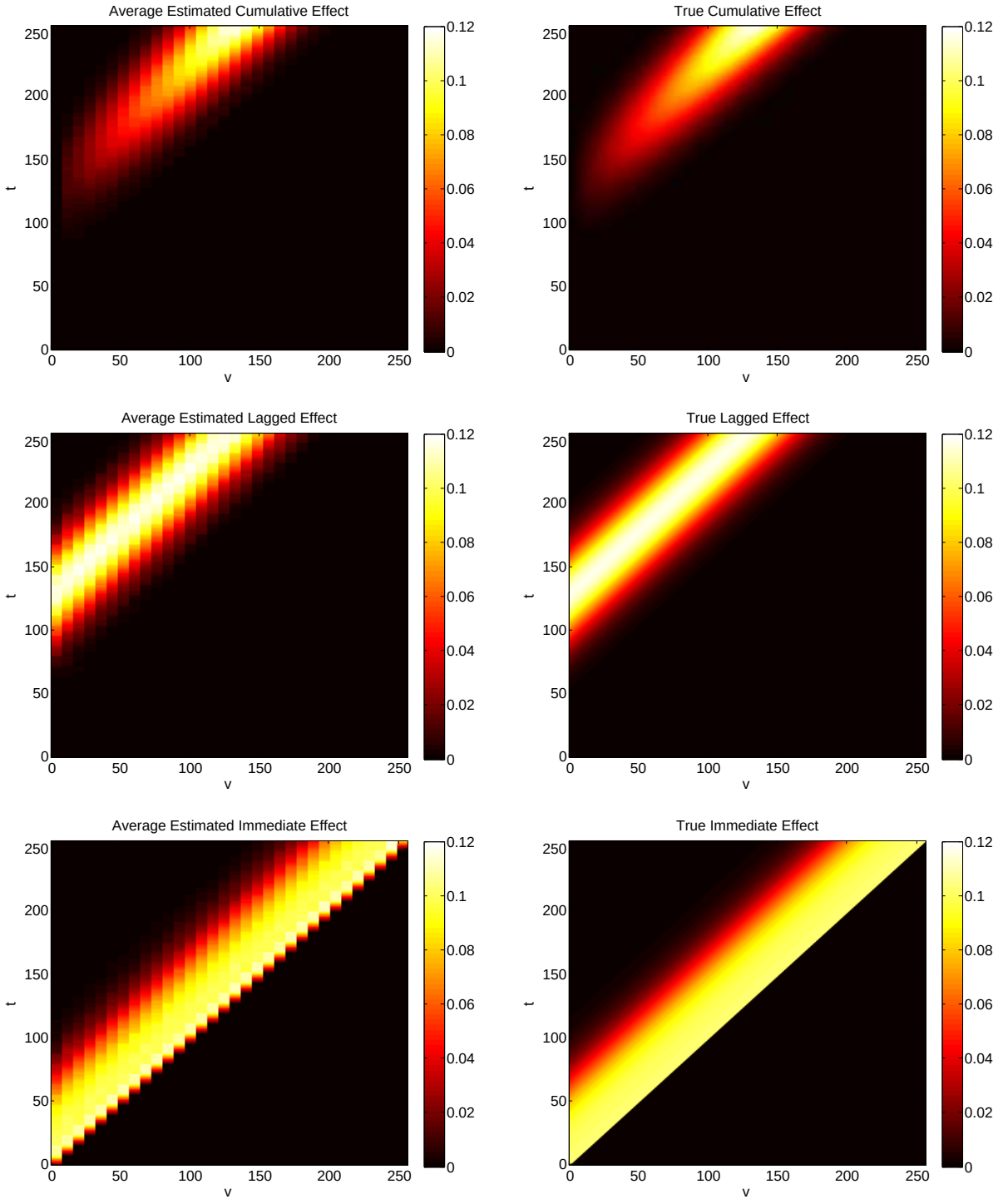


Figure 2.4: Left column: heat maps of average estimated $\beta(v,t)$ plotted as functions of t and v based on a sample size of $n = 45$ with $N = 150$ total curves. Right column: heat maps of the true $\beta(v,t)$ functions plotted as functions of t and v.

does not vary as the exposure is prolonged. The lagged effect represents a scenario where the effect of exposure is not dependent on amount of time exposed but the effect is lagged. The cumulative effect is a scenario in which the effect of exposure is not only lagged but the longer a subject is exposed, the greater the effect on outcome becomes.

The repeated measures simulation used a sample size of $n = 45$ with each subject contributing one to four curves for a total of $N = 150$ observations. In each simulated data set, 8.9% of subjects contributed only one trip, 13.3% contributed two trips, another 13.3% contributed three trips, and the majority, 64.5%, contributed four trips. These values were based off the Saint Louis Bus Study. The large sample simulation used a sample of $n = N = 1000$ with no repeated measures.

To evaluate inference, we propose the use of three metrics used in Meyer et al. (2014), false discovery rate (FDR), sensitivity (SEN), and experiment-wise error rate (EWER). First we define a flagged location (v, t) as one either belonging to ψ or having a SimBa Score less than 0.05. Now define FDR$_\epsilon$ as the number of flagged locations (v, t) with true value $\leq \epsilon$ divided by the total number of flagged locations. Next let SEN$_\Upsilon$ be the number of flagged locations (v, t) with true magnitude $> \Upsilon$ divided by the total number of locations with true magnitude $> \Upsilon$. Finally, EWER is calculated as the proportion of simulated datasets with at least one falsely discovered location. For the BFDR, we select a δ-intensity change of 0.05 as it is half the maximum signal of the true scenarios.

Figure 2.4 contains the true surfaces in the right column and the average of estimated surfaces across 200 hundred data sets based on the repeated measures simulation in the left column. Both the estimate of the cumulative effect and lagged effect maintain the integrity of the constraint while capturing the true effects. With the immediate effect we see the constraint isn't fully maintained, but the edge effects are minimal. On the whole, the average estimated surfaces effectively capture the magnitude and shape of the true surfaces. The estimated surfaces perform similarly as sample size increases which can be found in the Appendix.

Table 2.1: FDR, sensitivity and, experiment-wise error rate values by inference procedure. BFDR was calculated with a δ of 0.05 and a global α of 0.05.

Measure	Surface	BFDR	SimBaS
$FDR_{0.01}$	Cumulative	0.00%	0.02%
	Lagged	0.43%	0.00%
	Immediate	0.00%	0.00%
$FDR_{0.05}$	Cumulative	7.69%	16.1%
	Lagged	8.40%	0.12%
	Immediate	6.31%	0.21%
$SEN_{0.05}$	Cumulative	93.5%	92.0%
	Lagged	91.8%	8.88%
	Immediate	94.8%	38.6%
$SEN_{0.075}$	Cumulative	100%	99.7%
	Lagged	99.3%	12.6%
	Immediate	97.3%	46.1%
$EWER_{0.01}$	Cumulative	0.00%	0.04%
	Lagged	2.00%	0.00%
	Immediate	6.00%	1.50%

In Table 2.1 we examine inference in simulation where we see that both BFDR and SimBaS control FDR well for $\epsilon = 0.01$ for all scenarios. When ϵ is increased to 0.05, FDR increases for both SimBaS and BFDR though for the Lagged and Immediate scenarios, FDR is still small. For sensitivity, BFDR performs well for all scenarios and both levels of Υ. However the sensitivity of SimBaS only performs well for the cumulative scenario. The GBPV performs well under all scenarios rejecting the null of $\beta(v,t) = 0 \; \forall \; t \in \mathcal{T}, v \in \mathcal{V}$ for all datasets.

2.5 Example: Boilermaker Study

Magari et al. (2001) examined the effects of $PM_{2.5}$ on HRV as measured by SDNN from residual oil fly ash on apprentice boilermakers while Cavallari et al. (2008) describe a similar effect but in journeyman boilermakers. Additionally, Cavallari et al. (2008) compare both work effects and non-work day effects. To illustrate the usefulness of our method-

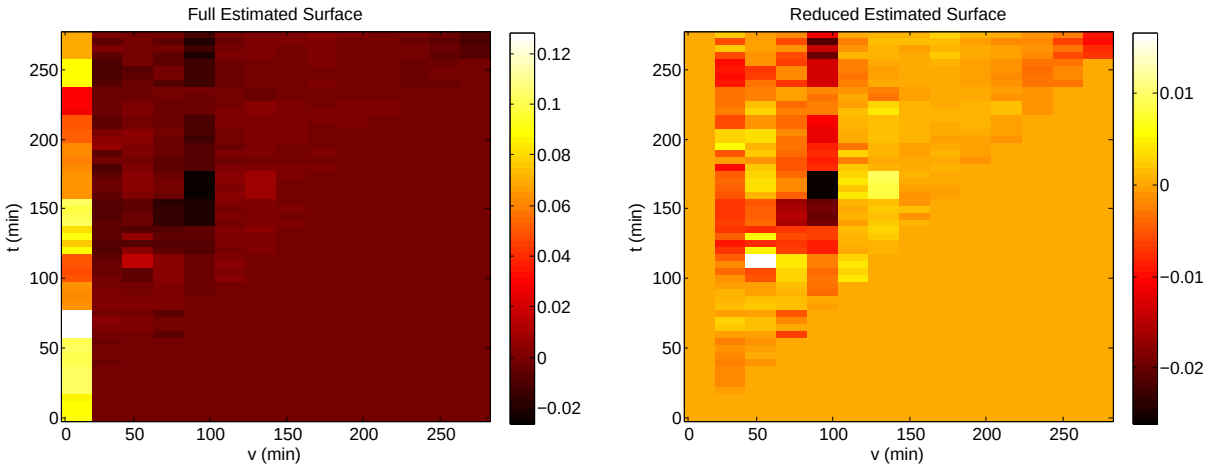


Figure 2.5: Left: the full estimated surface from the Boilermaker Study using $PM_{2.5}$ as the exposure and SDNN as the outcome. Right: estimated surface removing associations with minimal to no preceding exposure. Time scale is in minutes since the start of measurement. Both exposure and outcome were log-transformed prior to modeling.

ology, we propose jointly examining the data from these two studies. The resulting data consists of 31 subjects with five-minute average SDNN and $PM_{2.5}$ levels taken for 280 minutes, roughly four hours, resulting in 56 time points. As $PM_{2.5}$ is a natural exposure, our model is an appropriate choice for attempting to characterize the time-varying association of SDNN and $PM_{2.5}$. However, we must consider that associations early in the exposure time lack sufficient data. Thus in analysis, we ignore estimation of associations for the first four measurements in exposure. Consistent with the analysis performed in Harezlak et al. (2007), we use the $\log$ of each variable in modeling. Figure 2.5 contains results of the application of our method to the joint data for both including and excluding associations for the first four exposure measurements.

From the reduced estimated surface, we see that early in the exposure time, v, and early in the outcome time t, there is a positive association that quickly dissipates becoming a negative association to no association as time increases in both directions. The strongest negative effects occur between 50 and 100 minutes of exposure but are lagged affecting SDNN only after 100 minutes of measurement. Looking at Figure 2.5, we see that around 100 minutes there is a lull in exposure. This thus suggests that there is a possible delay

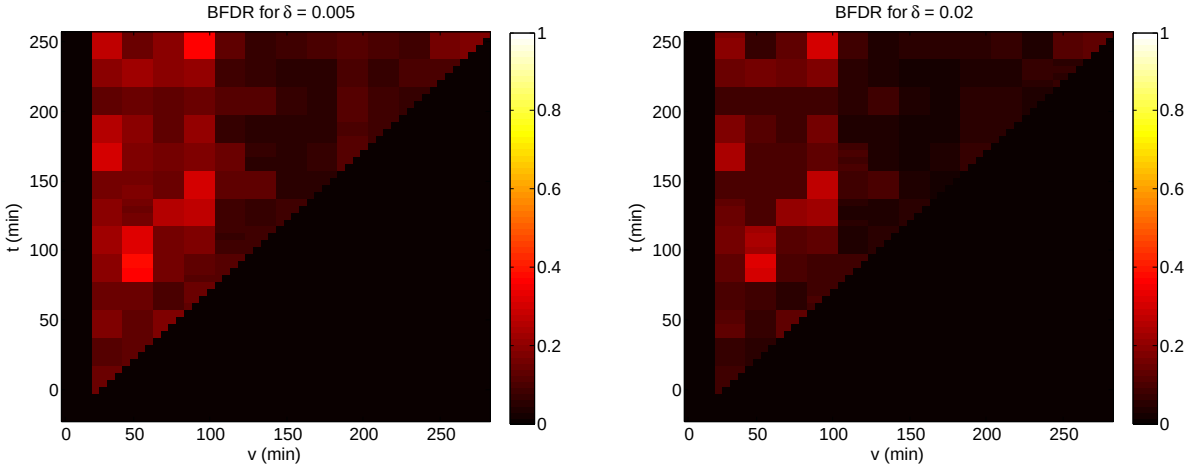


Figure 2.6: Posterior probabilities of the coefficients of the estimated surface for two levels of δ. The left contains the heat map of a low δ intensity change of 0.005 while the right heat map contains a high δ intensity change of 0.02, high and low relative the surface.

in the effect of of $PM_{2.5}$ on HRV. After around 100 minutes of exposure, the association is around zero but is still diminished from the higher levels of SDNN seen at the very beginning of the study. Interestingly, at the tail end of the study, we see another dip in the surface suggesting a negative relationship which appears to correspond to several spikes in PM, or falls just shortly there after.

In Figure 2.6, we see heat heat maps of the posterior probability that $|\beta(v,t)|$ is greater than a δ intensity change. For this analysis, we consider two levels of δ, one low and one high relative the effect sizes found in the surface. We select $\delta = 0.005$ for the low change and a $\delta = 0.02$ for the high. We can see that in both cases, probabilities do not exceed 0.5 suggesting that, at least at these levels of δ, there are no significant coefficients. In fact, no coefficients are flagged by the procedure at global α-level of 0.05 and increasing α to 0.1 similarly results in no flagged coefficients. Further, SimBaS was unable to detect an coefficients significantly different from zero, thus the exclusion of a heat map of those results. The GBPV also fails to reject at $\alpha = 0.05$ as the minimum SimBa Score is 0.5.

One possible reason for the lack of significant effects is that the data are a combination of journeyman and apprentices which may have different patterns of exposure thus poten-

tially muddling inference. Unfortunately, separate analyses of these two data sets were limited due to small sample size and thus not examined here. Still, the estimated surface behaves in a manner that is consistent with what both Magari et al. (2001) and Cavallari et al. (2008) saw in their analyses which did not joint model SDNN and $PM_{2.5}$ across time.

2.6 Discussion

Natural exposures commonly arise in the study of environmental health effects. Care is needed in modeling such data to ensure that different types of effects can be detected and appropriately modeled. Previous work on the HFLM has focused on overly smoothed models for exposure which, as demonstrated by Figure 2.1, can be quite spiky and unpredictable. Further, the potential for multiple curves sampled on the same individual has been ignored despite this being commonplace in environmental literate. Thus to accommodate the needs of increasingly complex data structures, a methodology capable of handling irregular data and repeated measures is warranted.

Here we've developed a wavelet-based historical functional linear model for repeated measures using a Bayesian modeling approach. To accomplish this, we proposed a novel use of wavelet-packets demonstrating their superiority in maintaining the historical constraint over regular wavelets. Previous work on HFLMs have ignored inference procedures. Here we adapt those explored by Meyer et al. (2014) to the historical setting. In simulation, we demonstrate the abilities of our model to correctly estimate several scenarios of interest to environmental health researchers. Further we show the benefits of the BFDR procedure for use as a tool for inference in the HFLM. Under certain circumstances, we show that SimBaS also performs well. In the boilermaker example, we show how our model can be applied to data consisting of an environmental exposure and measure of HRV. While we were unable to detect any significant differences, we were able to characterize the shape of the association in a manner consistent with previous studies.

Several aspects of the Historical Functional Mixed Model will require additional exploration. An alternative procedure to SimBaS, for instance, may be of interest to provide a better test of the null $\beta(v,t) = 0$ at a specific location. Additionally, the current methodology works best for Haar wavelets as they maintain the constraint better than other wavelets. However this results in a rather blocky, un-smoothed estimated surface. One potential solution comes from Wand and Omerod (2011) who examine penalized wavelet transformations. Such an approach may allows us to continue to use the Haar wavelets to main the constraint while smoothing away from it. Another direction is to exploit the function-on-function regression by Meyer et al. (2014) and implement a projection procedure in the spirit of Dunson and Neelon (2003) to project the unconstrained coefficients into the constrained space formed by the triangular basis functions used by Malfait and Ramsay (2003) and Harezlak et al. (2007).

Acknowledgements

This work was supported by grants from the National Institutes of Health (ES007142, ES000002, ES016454, CA134294).

References

Breysse, P. N., Delfino, R. J., Dominici, F., Elder, A. C. P., Frampton, M. W., Froines, J. R., Geyh, A. S., Godleski, J. J., Gold, D. R., Hopke, P. K., et al. (2013). US EPA particulate matter research centers: summary of research results for 2005-2011. *Air Quality Atmosphere and Health* **6,** 333–355.

Cavallari, J. M., Fang, S. C., Eisen, E. A., Schwartz, J., Hauser, R., and Herrick, R. F. (2008). Time Course of Heart Rate Variability Decline Following Particulate Matter Exposures in an Occupational Cohort. *Inhalation Toxicology* **20,** 415–422.

Donoho, D. L. and Johnstone, I. M. (1994). Ideal spatial adaptation by wavelet shrinkage. *Biometrika* **81,** 425–455.

Dubowsky, S. D., Suh, H., Schwartz, J., Coull, B. A., and Gold, D. R. (2006). Diabetes, obesity, and hypertension may enhance associations between air pollution and markers of systemic inflammation. *Environmental Health Perspectives* **114,** 992–998.

Dunson, D. B. and Neelon, B. (2003). Bayesian Inference on Order-Constrained Parameters in Generalized Linear Models. *Biometrics* **59,** 286–295.

Harezlak, J., Coull, B. A., Laird, N. M., Magari, S. R., and Christiani, D. C. (2007). Penalized solutions to functional regression problems. *Computational Statistics & Data Analysis* **51,** 4911–4925.

Huang, W., Zhu, T., Pan, X. C., Hu, M., Lu, S. E., Lin, Y., Wang, T., Zhang, Y. H., and Tang, X. Y. (2012). Air Pollution and Autonomic and Vascular Dysfunction in Patients

With Cardiovascular Disease: Interactions of Systemic Inflammation, Overweight, and Gender. *American Journal of Epidemiology* **176,** 117–126.

Huttunen, K., Siponen, T., Salonen, I., Yli-Tuomi, T., Aurela, M., Dufva, H., Hillamo, R., Linkola, E., Pekkanen, J., and Pennanen, A. (2012). Low-level exposure to ambient particulate matter is associated with systemic inflammation in ischemic heart disease patients. *Environmental Research* **116,** 44–51.

Ivanescu, A.E., Staicu, A.-M., Greven, S., Scheipl, F., and Crainiceanu, C. M. (2012). Penalized function-on-function regression. *Johns Hopkins University Working Paper*.

Kim, K., Şentürk, D., and Li, R. (2011). Recent history functional linear models for sparse longitudinal data. *Journal of Statistical Planning and Inference* **141,** 1554–1566.

Magari, S. R., Hauser, R., Schwartz, J., Williams, P. L., Smith, T. J., and Christiani, D. C. (2001). Association of Heart Rate Variability With Occupational and Environmental Exposure to Particulate Air Pollution. *Circulation* **104,** 986–991.

Malfait, N. and Ramsay, J. O. (2003). The historical functional linear model. *The Canadian Journal of Statistics* **31,** 115–128.

Malloy, E. J., Morris, J. S., Adar, S. D., Suh, H., Gold, D. R., and Coull, B. A. (2010). Wavelet-based functional linear mixed models: an application to measurement error-corrected distributed lag models. *Biostatistics* **11,** 432–452.

Meyer, M. J., Coull, B. A., Versace, F., and Morris, J. S. (2014). Bayesian Function-on-Function Regression for Multi-Level Functional Data. *In Preparation*.

Misiti, M., Misiti, Y., Oppenheim, G., and Poggi, J.-M. (2007).*Wavelets and their Applications*. ISTE Ltd.

Morris, J. S., Brown, P. J., Herrick, R. C., Baggerly, K. A., and Coombes, K. R. (2008). Bayesian Analysis of Mass Spectrometry Proteomic Data Using Wavelet-Based Functional Mixed Models. *Biometrics* **64,** 479–489.

Morris, J. S. and Carroll, R. J. (2006). Wavelet-based functional mixed models. *Journal of the Royal Statistical Society, Series B* **68,** 179–199.

Nason, G. P. (2008). *Wavelet Methods in Statistics with R*. Springer.

Neophytou, A. M., Hart, J. E., Cavallari, J. M., Smith, T. J., Dockery, D. W., Coull, B. A., Garshick, E., and Laden, F. (2013). Traffic-related exposures and biomarkers of systemic inflammation, endothelial activation and oxidative stress: a panel study in the US trucking industry. *Environmental Health* **12**.

Percival, D. B. and Walden, A. T. (2000). *Wavelet Methods for Time Series Analysis*. Cambridge University Press.

Ruppert, D., Wand, M. P., and Carroll, R. J. (2003). *Semiparametric Regression*. Cambridge University Press.

Scheipl, F. and Greven, S. (2012). Identifiability in penalized function-on-function regression models. *University of Munich Technical Report* **125.**

Scheipl, F., Staicu, A.-M., and Greven, S. (2014). Functional Additive Mixed Models. *Journal of Computational and Graphical Statistics, to appear*. Available at *arXiv:1207.5947v5*.

Wand, M. P. and Ormerod, J. T. (2011). Penalized wavelets: Embedding wavelets into semiparametric regression. *Electronic Journal of Statistics* **5,** 1654–1717.

3. Ordinal Probit Wavelet-based Functional Models for eQTL Analysis

Mark J. Meyer[1], Christoph Lange[1], Craig P. Hersh[2], and Brent A. Coull[1]

[1]Department of Biostatistics, Harvard School of Public Health
[2]Channing Division of Network Medicine, Brigham and Womens Hospital, Harvard Medical School

3.1 Introduction

Morris and Carroll (2006) introduce Wavelet-based Functional Mixed Models (WFMM) for function-on-scalar regression via Bayesian estimation. The WFMM is a flexible framework for modeling functional outcomes in a number of settings and indeed several authors have extended the methodology. Morris et al. (2008) introduce an inference procedure for the framework based on the Bayesian False Discovery Rate (BFDR) for mass spectrometry proteomic data. Malloy et al. (2010) develop the scalar-on-function regression analogue for repeated measures using wavelets for a functional covariate. Zhu, Brown, and Morris (2011, 2012) discuss robust adaptive regression and robust classification respectively. Meyer et al. (2014a,b) introduce the function-on-function extension of the WFMM and the historical functional linear model analogue respectively. All of this additional work makes the assumption that the functional outcome comes from a Gaussian Process which does not hold when the observed functional outcome is categorical.

Outside the WFMM framework, the existing functional literature on generalized outcomes deals with a scalar outcome regressed on a functional covariate. This model, the generalized functional linear model (GFLM), was developed for the linear, logistic, censored, and Poisson cases by James (2002) and Müller and Stadtmüller (2005). Further work by numerous authors since includes but is not limited to a penalized likelihood approach to modeling, the inclusion of single-index interactions, and an extension to the functional generalized additive model (Cardot and Sarda, 2005; Li, Wang, and Carroll, 2010; McLean et al., 2014). One area where the literature is lacking is the generalized function-on-scalar regression setting where the functional outcome is categorical in nature.

Such a setting is common when expression quantitative trait loci (eQTL) analysis is of interest. eQTL analysis examines the association between expression levels of a gene as measured by microarray probe sets and a fine mapping of single nucleotide polymorphisms (SNPs) to measure genotype in the same region as the gene. While expression

levels are normalized and thus gaussian, non-imputed SNPs take on the integer values $0, 1$ and 2. Combining this with the fine sampling or mapping and genetic order as determined by location on the chromosome allows us to consider SNP data a function of time. Thus a subject's curve is their ordered genotype with genetic order serving as a proxy for time, the more traditional measure in functional data. Multiple expression probe sets are often taken over the same gene. Researchers are interested in using eQTL analysis to determine which SNPs in the region are significantly associated with expression levels while adjusting for phenotypes and genetic ancestry.

Currently, the standard analysis for performing eQTL analysis is a pair-wise regression approach comparing an expression probe set to a single SNP at a time and determining a p-value. This procedure is repeated for each SNP sampled in the region. Ad-hoc adjustments are used including eQTL false discovery rate (FDR) and even Bonferroni Correction. Such an analysis can be seen, for example, in both Qiu et al. (2011) and Castaldi et al. (2014) where general linear models are used to assess association and adjust for phenotypes and population stratification. Some existing statistical literature proposes a joint modeling procedure treating genotype as a non-functional set of covariates and using a modified BIC approach to eliminate non-significant SNPs (Zak-Szatkowska and Bogdan, 2011; Frommlet et al., 2012). However the modified BIC approach does not allow for multiple probe sets. Flutre et al. (2013) propose a Bayesian model averaging (BMA) framework for a joint eQTL analysis across multiple probe sets from different tissues. But the BMA method only examines one candidate SNP at a time. Thus issue of running many models for all SNPs of interest remains and the need exists for a method that jointly models all SNPs and probe sets of interest.

In this paper, we introduce the Ordinal Probit Wavelet-based Functional Model (OPWAVFM) for regressing a generalized functional outcome and scalar covariates using a Bayesian approach. The Bayesian setting allows for adaptive regularization of coefficients and thus smoothing across the functional form of the outcome. The OPWAVFM constitutes an extension of the WFMM framework to the non-Gaussian setting and, by

building off it, allows for the inclusion of a large number of covariates of interest as well as a large number of outcome measurements. We propose a Markov Chain Monte Carlo (MCMC) algorithm for generating posterior estimates of model parameters. The algorithm combines the standard Bayesian Probit procedure with the WFMM framework which is achieved in part due to the latent variable representation of the Probit model.

As in Morris et al. (2008), we aim to exploit the flexibility of the WFMM framework in modeling the effects of multiple factors simultaneously on nonparametric fixed effects functions. Because the SNP data that motivates this research is ordinal in nature, we implement an ordinal regression which can be extended beyond three levels. Previous work in the WFMM context implements both the BFDR, Morris et al. (2008) and Malloy et al. (2010), and Simultaneous Band Scores (SimBaS), Meyer et al. (2014a,b). Thus we formulate the OPWAVFM version of these posterior inference procedures while noting that any statistic of interest can be calculated from our posterior estimates. Finally, we propose an extension of the OPWAVFM to the function-on-function regression setting which can also be considered an extension of Meyer et al. (2014a) to generalized outcomes.

The remainder of the paper is organized as follows: Section 3.2 presents the model formulation, MCMC algorithm, and inference procedures we've developed for implementing the OPWAVFM. Section 3.3 details the extension of the OPWAVFM and Meyer et al. (2014a) to the generalized function-on-function regression setting. In Section 3.4, we describe a simulation study demonstrating the abilities of our method. In Section 3.5, we apply the OPWAVFM to an example genomic data set and in Section 3.6 we give a discussion of the methodology.

3.2 Ordinal Probit Functional Model

Here we detail the modeling framework for the Ordinal Probit Wavelet-based Functional Model (OPWAVFM). For subjects $i = 1, \ldots, N$, assume we observe the categorical func-

tional variable $Y_i(t)$ on the grid $t = 1, \ldots, T$ where t indexes measurement occurrence and $Y_i(t)$ takes on the values $g = \{0, 1, 2\}$. Further assume we observe a scalar covariate or potential a set of scalar covariates which we denote as X. The values of X may be continuous or categorical. For the sake of derivation, we focus on the case where X is a single continuous scalar covariate as is the case in the motivating data.

Now suppose $Y_i(t)$ is actually the observable values of some latent process $Y_i^*(t)$. Then the behavior of $Y_i(t)$ is dictated by the relationship

$$Y_i(t) = \begin{cases} 0 & \text{if } Y_i^*(t) < c_1 \\ 1 & \text{if } c_1 \leq Y_i^*(t) < c_2 \\ 2 & \text{if } Y_i^*(t) \geq c_2 \end{cases} \tag{3.1}$$

for cut points c_1 and c_2, satisfying $c_1 < c_2$. This formulation can be extended to more levels, however we restrict it to the three level case. Using the mapping in Model (3.1), the probability that $Y_i(t)$ equals the gth level can be expressed as

$$P(Y_i(t) = g) = P(Y_i^*(t) \in (c_g, c_{g+1})),\ g = 0, 1, 2 \tag{3.2}$$

where c_0 and c_3 will vary depending on the support of $Y_i^*(t)$. Let the form of $Y_i^*(t)$ be the function-on-scalar regression model in the spirit of Morris and Carroll (2006)

$$Y_i^*(t) = X_i\beta(t) + E_i(t) \tag{3.3}$$

where the model errors $E_i(t)$ could come from a variety of distributions.

If we assume Gaussian Process errors, then $Y_i^*(t)$ is also Gaussian. Further, the probability function, $P(\cdot)$, can now be defined as the cumulative distribution function (CDF) from a Gaussian. Normalizing $Y_i^*(t)$, we can re-express Model (3.2) in terms of the the CDF of the standard Gaussian, denoted $\Phi(\cdot)$. The probabilities at a fixed t for subject i are then given by

$$P(Y_i(t) = g) = \Phi\left(c_{g+1} - X_i\beta(t)\right) - \Phi\left(c_g - X_i\beta(t)\right) \tag{3.4}$$

where $g = 0, 1, 2$. This Gaussian assumptions results in the Probit formulation of the model.

Models (3.3) and (3.4) are formulated for continuous functions. However we only observe discretized realizations of each function. Thus the discretized form of Model (3.3) is

$$\mathbf{Y}^* = X\boldsymbol{\beta} + \mathbf{E},\ \mathbf{E} \sim \mathcal{GP}(0, \Sigma_{\mathbf{E}}) \tag{3.5}$$

where $\mathbf{Y}^*$ and $\mathbf{E}$ are $N \times T$, X is $N \times 1$, and $\boldsymbol{\beta}$ is $1 \times T$. The covariate matrix, X, can be of size $N \times P$ depending on the desired number of covariates. If $P > 1$, then $\boldsymbol{\beta}$ becomes $P \times T$ with one function per covariate. Now let y_{it} represent subject i's tth outcome, β_t correspond to the tth element of $\boldsymbol{\beta}$, and x_i denote subject i's covariate pattern. Model (3.4) can then be written as $P(y_{it} = g) = \Phi\left(c_{g+1} - x_i\beta_t\right) - \Phi\left(c_g - x_i\beta_t\right)$ for $g = 0, 1, 2$.

Non-functional Bayesian Probit procedures utilize a similar latent variable formulation to to produce model estimates sampling from the latent outcome. In this approach, the latent outcome is assumed independent which poses an issue as we extend this idea to the functional setting: the columns of $\mathbf{Y}^*$ are not necessarily independent. However, using the function-on-scalar formulation implemented by Morris and Carroll (2006), we can assume independence after wavelet transforming Y^* and model in the transformed space.

Note the observed measurements do not necessarily have to have a time element, just a natural ordering. Further, measurement occurrences do not need be equally spaced though we assume we observe a measurement for each subject at each t. The methodology can also be extended to more levels, but for now we restrict our derivation to the three-level case which corresponds to the motivating genetic data. Also, for identifiability, we do not specify Model (3.5) with an intercept. That is, X only contains covariates of interest as the intercept function is assumed to equal zero.

3.2.1 Wavelet-based Modeling of the Latent Outcome

Morris and Carroll (2006) develop a function-on-scalar regression model for hierarchical data. Thus their formulation allows for the observation and modeling of multiple curves measured on each subject. When only one curve is observed per subject, their model

essentially reduces to the latent variable model found in Model (3.3). Given this relationship, we can use their modeling framework, alongside a variation of Bayesian Probit regression, to get estimates of $\beta(t)$.

Model Formulation

Let $y_i^*(t) = \sum_{j=1}^{T^*} y_{ij}^{*W} \xi_j(t) \Delta_{tj}$ be the Karhunen-Loève expansion of the functional latent response where $\Delta_{tj} = t_j - t_{j-1}$. For convenience, we assume measurements are equally spaced and thus $\Delta_{tj} = 1$. The basis expansion for the latent outcome, $\xi_j(t)$, requires only that independence be induced or at least reasonably assumed post decomposition. Previous authors have implemented a variety of expansions. Consistent with Morris and Carroll (2006) and Meyer et al. (2014a), we use wavelets to model the latent response.

Working from the discretized model, Model (3.5), and applying a Discrete Wavelet Transformation (DWT) the latent outcome gives the decomposition to $\mathbf{Y}^* = \mathbf{Y}^{*W} W_Y$ where $\mathbf{Y}^{*W}$ are the resulting wavelet-space coefficients and W_Y is a matrix of wavelet basis functions. This decomposition is the matrix form of the Karhunen-Loève expansions previously described. Decomposing $\boldsymbol{\beta}$ gives $\boldsymbol{\beta} = \boldsymbol{\beta}^W W_Y$ and the DWT applied to $\mathbf{E}$ results in $\mathbf{E} = \mathbf{E}^W W_Y$. Given these decompositions, Model (3.5) can be expressed as

$$\mathbf{Y}^{*W} W_Y = X \boldsymbol{\beta}^W W_Y + \mathbf{E}^W W_Y. \tag{3.6}$$

Note that the matrix representations of the wavelet basis are orthogonal, thus $W_Y W_Y' = I_{T^*}$. Post-multiplying Model (3.6) by W_Y' gives

$$\mathbf{Y}^{*W} = X \boldsymbol{\beta}^W + \mathbf{E}^W \tag{3.7}$$

where $\mathbf{E}^W \sim \mathcal{GP}(0, \Sigma_{\mathbf{E}^W})$. We then use an MCMC procedure to obtain posterior estimates of $\boldsymbol{\beta}^W$, projecting them back into the data space before performing inference.

Prior Specification and Identifiability

As with Morris and Carroll (2006) and Meyer et al. (2014a), we assume independence in the wavelet space. Thus Model (3.7) can be split up into a series of T^* separate models for each coefficient in the Y-wavelet space, double-indexed by (j,k)

$$\mathbf{y}^{*W}_{(j,k)} = X\beta^W_{(j,k)} + \mathbf{E}^W_{(j,k)}. \tag{3.8}$$

Consistent with previous literature on wavelet-based regression (Morris and Carroll, 2006; Malloy et al., 2010; Meyer et al., 2014a,b), we place spike and slab priors on the coefficients $\beta^W_{(j,k)} = \left\{\beta^W_{(p,jk)}\right\}$ where p indexes the number of columns of X. Thus the prior on the coefficients from Model (3.8) is

$$\beta^W_{(j,k)} \sim \gamma_{(p,jk)}\mathcal{N}(0,\tau_{pj}) + (1-\gamma_{p,jk})d_0, \;\; \gamma_{(p,jk)} \sim \mathcal{B}(\pi_{pj}) \tag{3.9}$$

where $\mathcal{B}$ denotes the binomial distribution and d_0 is a point-mass distribution at zero. This adaptive regularization performs smoothing in the Y-wavelet space.

Regularization parameters can either be sampled or fixed. Both Morris and Carroll (2006) and Malloy et al. (2010) rely on an empirical Bayes approach to estimate and then fix τ_{pj} and π_{pj}. However Zhu, Brown, and Morris (2011) and Meyer et al. (2014a,b) all propose prior distributions using an inverse-gamma for τ_{pj} and a beta for π_{pj}. Thus we place priors on both of the form

$$\tau_{pj} \sim \mathcal{IG}(a_\tau, b_\tau) \text{ and } \pi_{pj} \sim Beta(a_\pi, b_\pi)$$

where hyper parameters a_τ, b_{tau}, a_π, and b_π are fixed and based on the empirical Bayes estimates found in Morris and Carroll (2006).

An assumption of the standard Bayesian Probit regression is that the mapping from $Y_i(t)$ to $Y_i^*(t)$ captures the location and scale of $Y_i(t)$. Thus certain parameters from the model are not estimated and are, in fact, not identifiable. As noted in the introduction to Section 3.2, we have not formulated a model with an intercept because the intercept in this

setting is not identifiable. Thus a priori we assume the intercept is fixed and equal to zero and so, without loss of generality, it is excluded from the model formulation. Similarly, the model error variance components of the latent variable in Model (3.5) are not identifiable and we thus assume $\Sigma_{\mathbf{E}}$ to be fixed. Our prior specification for the latent variable $\mathbf{Y}^*$ is then

$$\mathbf{Y}^*|X, \boldsymbol{\beta} \sim \mathcal{N}(X\boldsymbol{\beta}, \Sigma_{\mathbf{E}}) \tag{3.10}$$

where $\Sigma_{\mathbf{E}}$ is not only fixed but set to the $N \times N$ identity matrix, $\Sigma_{\mathbf{E}} = I_N$. Note the prior specification for the latent variable is not in the wavelet space. Thus our MCMC algorithm will require performing the DWT and its inverse at every step.

The last identifiability concern regards the cut points dictating the mapping of $Y_i(t)$ to $Y_i^*(t)$, i.e. c_1 and c_2. Two interrelated issues arise with attempting to sample these parameters. The first is ensuring the constraint, $c_1 < c_2$, is enforced. The second, which is related to the first, is that we can't estimate both simultaneously. Several approaches exist to both ensure estimability and enforce the constraint. The first approach involves first sampling c_1 from a mean zero normal distribution with fixed variance and then sampling c_2 as $c_1 + \delta$ where δ comes from an exponential distribution with fixed rate. This approach is independent of the latent variable and thus also of the data which is problematic. Further, though the constraint is satisfied, identifiability is still a concern since the approach tries to estimate both values. A second approach fixes c_1, usually at zero, and then samples c_2 from a uniform distribution, $\mathcal{U}(a_{c_2}, b_{c_2})$, where a_{c_2} and b_{c_2} are determined using both c_1 and information from the data.

3.2.2 MCMC Algorithm

Given the prior specifications in Section 3.2.1, we now describe the MCMC algorithm for obtaining estimates of $\beta(t)$. The procedure is a modification of the standard Bayesian Probit regression modified to accommodate an ordinal and functional outcome and scalar predictor or predictors. The standard algorithm can vary slightly in order depending on

approach to estimating the cut points. We begin our procedure by first sampling latent variable, then updating the cut points, and finally sampling the parameters from the latent variable model. Two additional steps are needed to project the latent variable into the wavelet space and to bring the coefficients into the data space.

Define $\mathbf{y}$ as the vectorized form of $\mathbf{Y}$, the matrix of ordinal outcomes and likewise define $\mathbf{y}^*$ as the vectorized form of $\mathbf{Y}^*$, the matrix form of the latent variable. Further let $\boldsymbol{\mu}^* = \mathbb{E}(\mathbf{y}^*)$, the expected value of the vectorized latent variable. Our MCMC algorithm is then

Step 1: Update the latent variable $\mathbf{y}^*|\mathbf{y}, \boldsymbol{\mu}^*, c_1, c_2$ using truncated normals of the form

$$\begin{aligned}
\mathbf{y}^*|\mathbf{y} = 0, \boldsymbol{\mu}^*, c_1 &\sim \mathcal{N}(\boldsymbol{\mu}^*, I)1(\mathbf{y}^* \in (-\infty, c_1)) \\
\mathbf{y}^*|\mathbf{y} = 1, \boldsymbol{\mu}^*, c_1, c_2 &\sim \mathcal{N}(\boldsymbol{\mu}^*, I)1(\mathbf{y}^* \in (c_1, c_2)) \\
\mathbf{y}^*|\mathbf{y} = 2, \boldsymbol{\mu}^*, c_2 &\sim \mathcal{N}(\boldsymbol{\mu}^*, I)1(\mathbf{y}^* \in (c_2, \infty))
\end{aligned}$$

where I denotes the identity matrix and $1(\cdot)$ is the indicator function. Note too that $\boldsymbol{\mu}^*$ is taken to be $X\boldsymbol{\beta}^{W(m)}$ where $\boldsymbol{\beta}^{W(m)}$ is the mth posterior draw of $\boldsymbol{\beta}^W$.

Step 2: Update the cut point $c_2|\mathbf{Y}^*, \mathbf{Y}, c_1$ using the uniform distribution

$$c_2|\mathbf{y}^*, \mathbf{y}, c_1 \sim \mathcal{U}(a, b)$$

for $a = \max\left[\max(\mathbf{y}^*|\mathbf{y} = 1), c_1\right]$ and $b = \min(\mathbf{y}^*|\mathbf{y} = 2)$.

Step 3: Perform DWT on latent variable: $\mathbf{Y}^* \overset{\text{DWT}}{\longrightarrow} \mathbf{Y}^* W_Y$

Step 4: Update $\boldsymbol{\beta}^W$ using

$$\beta^W_{(p,jk)}|\mathbf{y}^W_{(j,k)}, \beta^W_{(-p),jk}, \boldsymbol{\Sigma}^W \sim \gamma_{p,jk}\mathcal{N}(\mu_{p,jk}, \varepsilon_{p,jk}) + (1 - \gamma_{p,jk})d_0$$

Where the mixture probability $\alpha_{p,jk}$ is given by

$$\alpha_{p,jk} = \Pr\left(\gamma_{p,jk} = 1|\mathbf{y}^*_{(j,k)}, \beta^W_{(-p),jk}, \boldsymbol{\Sigma}^*\right) = O_{p,jk}/\left(O_{p,jk} + 1\right)$$

for $O_{p,jk} = \pi_{pj}/(1 - \pi_{pj})\text{BF}_{p,jk}$ and $\text{BF}_{p,jk} = (1 + \tau_{p,jk}/V_{p,jk})^{-1/2} \exp\left\{\frac{1}{2}\zeta^2_{p,jk}(1 + V_{p,jk}/\tau_{p,jk})\right\}$ and $\mu_{p,jk} = \hat{\beta}^W_{(p,jk),\text{MLE}}(1 + V_{p,jk}/\tau_{p,jk})^{-1}$ and $\varepsilon_{p,jk} = V_{p,jk}(1 + V_{p,jk}/\tau_{p,jk})^{-1}$. Both $\hat{\beta}^W_{(p,jk),\text{MLE}}$

and $V_{p,jk}$ are initial values taken from a maximum likelihood estimation of the latent variable model.

Step 5: Update τ_{pj} and π_{pj} using

$$\tau_{pj}|a_\tau, b_\tau, \gamma_{p,jk}, \beta^W_{(p,jk)} \sim \mathcal{IG}\left(a_\tau + \frac{1}{2}\gamma_{p,jk}, b_\tau + \frac{1}{2}\gamma_{p,jk}\left\{\beta^W_{(p,jk)}\right\}^2\right)$$
$$\pi_{pj}|a_\pi, b_\pi, \gamma_{p,jk} \sim Beta\left(a_\pi + \gamma_{p,jk}, b_\pi + \gamma_{p,jk}\right)$$

for a_τ, b_τ, a_π, and b_π fixed.

Step 6: Project $\boldsymbol{\beta}^W$ into the data space using appropriate inverse DWT matrix: $\boldsymbol{\beta} = \boldsymbol{\beta}^P W_Y$. Because the algorithm involves the projection of $\boldsymbol{\beta}^P$ back into the data space, post processing only involves the calculation of summary measures based on the posterior samples and statistics to perform inference.

3.2.3 Posterior Functional Inference

We wish to perform inference on the surface of associations $\beta(t)$. Previous authors have detailed two procedures for inference: the BFDR and SimBaS (Müller, Parmigiani, and Rice (2006); Morris et al. (2008); Malloy et al. (2010); Meyer et al. (2014a,b) and references therein). Both procedures have many advantages over existing approaches in the context of genotype and gene expression association studies. While existing methods do control for the false discovery rate, it's an ad-hoc adjustment while both BFDR and SimBaS adjust for multiplicity in their formulation. Further, our inference procedures are based off of a joint model as opposed to pair-wise comparisons. For completeness, we now discuss these procedures in the Bayesian OPWAVFM setting.

Our modeling procedure occurs on the inverse-Probit scale and as such we can perform inference on that scale. But our inference can be informed by the Probit scale which, as a probability, is restricted to being between 0 and 1. For $m = 1, \ldots, M$ MCMC samples, $\beta^{(m)}(t)$ is the mth draw from the posterior. Then for a specific $v,\ v = 1, \ldots, V$, and $t,\ t = 1, \ldots, T$, we calculate $Pr\left\{|\beta(t)| > \delta | y\right\} \approx \frac{1}{M}\sum_{m=1}^{M} 1\left\{\left|\beta^{(m)}(t)\right| > \delta\right\}$. Now δ can be

selected on either the inverse-probability scale or on the probability. If we choose on the probability-scale, we first select a desired change in probability, π^{δ}. Then the δ intensity change in the effect is given by $\delta = \Phi^{-1}(\pi^{\delta})$. To ensure δ is positive, we restrict π^{δ} such that $\pi^{\delta} \in (0.5, 1)$. The mapping is consistent with choosing a $\delta \in (0, \infty)$, but the choice is now more directed. Thus the BFDR can be reformulated as

$$P_{BFDR}(t) = Pr\left\{|\beta(t)| > \Phi^{-1}(\pi^{\delta})|y\right\} \approx \frac{1}{M}\sum_{m=1}^{M} 1\left\{\left|\beta^{(m)}(t)\right| > \Phi^{-1}(\pi^{\delta})\right\}.$$

Then for a pre-specified global FDR-bound α, we flag a set of locations satisfying $\psi = \{(t) : P_{BFDR}(t) \geq \nu_{\alpha}\}$ where $\nu_{\alpha} = P_{(\lambda)}$. And, given the ordered set $\left\{P_{(r)} : r = 1, \ldots, R\right\}$ for $R = VT$, the cutoff value $\lambda = \max\left[r^* : \frac{1}{r^*}\sum_{r=1}^{r^*}\left\{1 - P_{(r)}\right\} \leq \alpha\right]$. Previously, our choice of δ was somewhat arbitrary. However in this setting, we have a little more guidance in selecting the intensity change.

As discussed in Meyer et al. (2014a,b), the SimBaS procedure does not require the selection of an intensity change. The formulation of SimBaS for OPWAVFM is best done on the inverse-probability scale, i.e. on the $\beta(t)$ estimates themselves. The resulting scores can then be linked to the function transformed to the probability scale if desired. Working from the joint credible bands described by Ruppert, Wand, and Carroll (2003), we construct a $100(1-\alpha)\%$ credible band about $\beta(t)$. For the mean and standard deviation of $\beta(t)$ taken over the M MCMC samples, we construct the interval $I_{\alpha}(t) = \hat{\beta}(t) \pm q_{(1-\alpha)}\left[\widehat{\text{St.Dev}}\left\{\hat{\beta}(t)\right\}\right]$. The value $q_{(1-\alpha)}$ is then the $(1-\alpha)$ quantile taken over the M MCMC samples of the quantity

$$Z^{(m)} = \underset{v \in \mathcal{V}, t \in \mathcal{T}}{\max}\left|\frac{\beta^{(m)}(t) - \hat{\beta}(v,t)}{\widehat{\text{St.Dev}}\left\{\hat{\beta}(t)\right\}}\right|.$$

Next we vary the values of α, noting the minimum α at which $I_{\alpha}(t)$ excludes 0. These values, more formally defined by $P_{SimBaS}(t) = \min\{\alpha : 0 \notin I_{\alpha}(v,t)\}$, constitute the scores for SimBaS.

More directly, we can calculate SimBa Scores using

$$P_{SimBaS}(t) = \frac{1}{M}\sum_{m=1}^{M} 1\left\{\left|\frac{\hat{\beta}(t)}{\widehat{\text{St.Dev}}\left\{\hat{\beta}(t)\right\}}\right| \geq Z^{(m)}\right\}.$$

Each score essentially provides a test of the null $H_0 : \beta(t) = 0$ for a specific location t. On the probability scale, the scores corresponds to testing the null $H_0 : \Phi^{-1}(\beta(t)) = 0.5$. We can also calculate a global Bayesian p-value (GBPV) using SimBaS. Similar to Meyer et al. (2014a,b), the form of the GBPV is $P_{GBPV} = \min_t\{P_{MAPs}(t)\}$ which we can use to test the global null hypothesis $H_0 : \beta(t) = 0 \; \forall \; t = 1, \ldots, T$.

3.3 Extension to Generalized Function-on-Function Regression

Model (3.5) allows X be of size $N \times P$ which, in function-on-scalar regression, allows for any number of scalar covariates. In fact, we can let P get relatively large with respect to T, the total number of measurements observed for $Y_i(t)$. In other words, we can let X become $X(v)$ for the grid $v = 1, \ldots, V$ which is to say the formulation allows for a functional covariate. The values T and V need not be equal. Further, t and v only index measurement occurrence and thus do not necessarily have to represent the same grid. In other words, $Y_i(t)$ may be sampled more finely than $X_i(v)$ or vice versa. We now present the Ordinal Probit Wavelet-Based Function-on-Function Regression (OPWAVFFR) as an extension of the OPWAVFM.

Formulating the OPWAVFFR only requires minor alterations the model describe above. The latency assumption presented in Model (3.1) holds as does the probability described in Model (3.2). To incorporate subject i's functional covariate $X_i(v)$, we re-express Model (3.3) as

$$Y_i^*(t) = \int_{\mathcal{V}} X_i(v)\eta(v,t)dv + E_i(t) \tag{3.11}$$

where $\mathcal{V}$ is the support of $X_i(v)$. We are now interested in the estimation of the surface $\eta(v,t)$. Assuming the errors are Gaussian once again gives us the Probit model with probabilities for subject i at a fixed t given by

$$P(Y_i(t)=g)=\Phi\left(c_{g+1}-\int_{\mathcal{V}}X_i(v)\eta(v,t)dv\right)-\Phi\left(c_g-\int_{\mathcal{V}}X_i(v)\eta(v,t)dv\right)$$

for $g=0,1,2$. The discretized form of Model (3.11) is $\mathbf{Y}^*=\mathbf{X}\boldsymbol{\eta}+\mathbf{E}$ where $\mathbf{Y}^*$ and $\mathbf{E}$ are $N\times T$, $\mathbf{X}$ is $N\times V$, and $\boldsymbol{\eta}$ is $V\times T$.

The MCMC algorithm described in Section 3.2.2 can accommodate an X design matrix of reasonable size, however computational burden increases as P increases. Thus in the function-on-function model we must either limit the size of V or perform data reduction. Meyer et al. (2014a) explores a function-on-function regression for hierarchical data with a Gaussian Process outcome. They suggest the use of Wavelet-Principal Components (wPC) for decomposition of a functional covariate when data reduction is needed. The formulation of the latent variable in the OPWAVFFR then follows the procedure described in Meyer et al. (2014a) for the case where only a single set of curves, $\{Y_i(t),X_i(t)\}$ is observed on subject as opposed to multiple. Modeling of the OPWAVFFR can easily take place in the OPWAVFM context given a design matrix X containing not a set of scalar covariates, but measurements from a functional covariate projected into a desired space.

Denote the functional covariate as $\mathbf{X}$ and decompose it using a DWT, $\mathbf{X}=\mathbf{X}^WW_X$. Next decompose $\mathbf{X}^W$ using a singular value decomposition, $\mathbf{X}^W=\mathbf{X}^PP_{svd}$ where P_{svd} is the matrix of right singular vectors. The function-on-function representation of the wavelet-space model in Model (3.6)

$$\mathbf{Y}^{*W}W_Y=\mathbf{X}^WW_XP_{svd}P_{svd}'W_X'\boldsymbol{\eta}^PW_Y+\mathbf{E}^WW_Y \tag{3.12}$$

given the decomposition $\boldsymbol{\eta}=P_s'W_x'\boldsymbol{\eta}^PW_Y$. Note that since W_X and P_S are orthogonal, $W_XP_{svd}P_{svd}'W_X'=I_P$ where P is number of columns of $\mathbf{X}$. Thus after post-multiplying by W_Y', Model (3.12) reduces to $\mathbf{Y}^{*W}=\mathbf{X}^W\boldsymbol{\eta}^P+\mathbf{E}^W$ or essentially Model (3.7) with $\mathbf{X}^W$ replacing X and $\boldsymbol{\eta}^P$ replacing $\boldsymbol{\beta}$.

Alterations to Sections 3.2.2 and 3.2.3 are minor in order to implement the OPWAVFFR. For instance, the MCMC procedure remains the same with the exception of **Step 3** which now involves not only the inverse DWT for the Y wavelet space but also the inverse of the transformation used for $X_i(v)$. Additionally, instead of performing inference on $\beta(t)$, Section 3.2.3 can be modified to accommodate $\eta(v,t)$. For a detailed description of the formulation of the BFDR and SimBaS for a surface of coefficients, see Meyer et al. (2014a).

3.4 Simulation

We simulated data by generating probabilities for subject i at measurement t using

$$P_0 = \Phi\left(c_1 - X_i\beta(t)\right),\ P_1 = \Phi\left(c_2 - X_i\beta(t)\right) - \Phi\left(c_1 - X_i\beta(t)\right),$$
$$\text{and } P_2 = 1 - \Phi\left(c_2 - X_i\beta(t)\right)$$

where X_i is drawn from a standard normal distribution to mimic the normalized expression probe sets. We select $c_0 = 0$ and $c_2 = 0.5$. Values of $Y_i(t)$ were generated by first sampling a standard uniform random variable for each subject i and measurement t, call that value $u \sim \mathcal{U}(0,1)$, and then assigning 0, 1, or 2 using

$$Y_i(t) = 0 \iff u \in (0, P_0),\ Y_i(t) = 1 \iff u \in (P_0, P_0 + P_1)$$
$$\text{and } Y_i(t) = 2 \iff u \in (P_0 + P_1, 1).$$

The total number of measurements generated was $T = 256$. For the DWT on the latent outcome, a choice of padding is needed. Consistent with Malloy et al. (2010) and Meyer et al. (2014a,b), we use zero padding which in preliminary simulations demonstrated the least amount of edge effects. Other choices of padding, such as symmetric, can be implemented however they tend to amplify edge effects.

Two hundred datasets were generated for four scenarios of $\beta(t)$ for a single standard normal covariate. Two of these scenarios have different shaped peaks centered in a region

of interest:

$$\beta(t) = \frac{10}{\sqrt{2\pi(10)}} \exp\left[-\frac{1}{2}\frac{(t-128)^2}{10}\right], \text{ single normal}$$
$$\beta(t) = \frac{2}{5} \max\left[1 - \frac{|t-128|}{32}, 0\right], \text{ single triangle.}$$

The remaining scenarios have a central peak of the same magnitude as the single normal and the single triangle centered at $t = 128$ along with an additional attenuated peak centered at $t = 44$:

$$\beta(t) = \frac{10}{\sqrt{2\pi(10)}} \exp\left[-\frac{1}{2}\frac{(t-128)^2}{10}\right] + \frac{10}{3\sqrt{2\pi(10)}} \exp\left[-\frac{1}{2}\frac{(t-44)^2}{10}\right], \text{ double normal}$$
$$\beta(t) = \frac{2}{5} \max\left[1 - \frac{|t-128|}{32}, 0\right] + \frac{1}{7} \max\left[1 - \frac{|t-44|}{32}, 0\right], \text{ double triangle.}$$

Each of these true scenarios is included as the dark gray solid curve in Figure 3.1. After a burn-in of 5000 iterations, we sampled 1000 iterations for posterior estimation and inference. For each model, we use Daubechies wavelets with four vanishing moments and zero padding.

Because we wish to include multiple probe sets as covariates in our application, we also examine the abilities of the OPWAVFM to detect signals from multiple covariates. One potential issue with the inclusion of multiple probe sets in application is that it may induce collinearity. Thus our final simulation setting involves a combination of two scenarios: the single normal, denoted as $\beta_1(t)$, and the double normal, denoted as $\beta_2(t)$. To examine the effects of collinearity, we compare three levels of correlation between the two covariates: no correlation ($r = 0$), moderate correlation ($r = 0.5$), and high correlation ($r = 0.9$). The two covariates, x_1 and x_2, for this setting come from a mean zero bivariate normal distribution with $\text{var}(x_1) = \text{var}(x_2) = 1$ and correlation varying as described, $\text{corr}(x_1, x_2) = r = 0, 0.5$, and 0.9. To evaluate this scenario and the previous four single covariate settings, we examine the average point-wise bias calculated as the average of the difference between the estimate and the truth at each time point and root-Mean Square Error (rMSE) calculated for each simulated data set.

To evaluate inference, we implement two measures used by Meyer et al. (2014a,b) to assess both BFDR and SimBaS. First define a flagged coefficient as a coefficient belonging to the set ψ when using BFDR with global $\alpha = 0.05$ and score less than 0.05 when using SimBaS. Define the measure false discovery rate, denoted FDR_ε, as the number of flagged coefficients with true value $\leq \varepsilon$ divided by the total number of flagged locations. The second measure is sensitivity, denoted SEN_Υ, which we define as the number of flagged coefficients with true magnitude $> \Upsilon$ divided by the number of true coefficients with magnitude $> \Upsilon$. For our evaluation, we consider a range of both ε and Υ with $\varepsilon = 0.001, 0.002, 0.003, \ldots, 0.051$ and $\Upsilon = 0.05, 0.055, 0.01, \ldots, 0.3$. We compare these values for SimBaS as well as BFDR for $\pi_\delta = 0.501, 0.502, 0.503, \ldots, 0.6$ which, on the inverse-probability scale, corresponds to $\delta = 0.0025, 0.0050, 0.0075, \ldots, 0.2533$.

Figure 3.1 shows the average of the posterior estimates of each $\beta(t)$ along with the 95th percentile of estimates (in light gray). The true $\beta(t)$ is plotted for each scenario as well for comparison. In each scenario, we see the 95th percentile bands consistently include the truth. Additionally, the average of the posterior estimates effectively reproduces the true function. The OPWAVFM consistently detects the peaks in each scenario. While it struggles with the shape of the base of the peaks, particularly with the attenuated secondary peaks in the double scenarios, it's still able to detect them. The regions with the most issues are when the true functions transition to and from zero.

Such struggles are better demonstrated in the plot of the average bias found in Figure 3.2. At the transition points, we see the bias is negative suggesting underestimation for each scenario. At the largest peaks in all four scenarios, bias tends to be slightly overestimated. In the double peak scenarios, the double triangle is more biased on the smaller peak than the double normal. But those trouble spots aside, bias is small, mostly falling between 0.01 and -0.01 across t. All approaches are reasonably unbiased in regions where the truth equals or is close to 0. Figure 3.2 also contains box plots of rMSE for each scenario. The rMSE values confirm our assessment of minimal bias for each scenario. Not surprisingly, the more complicated the scenario, the larger rMSE tends to be. Though on the whole,

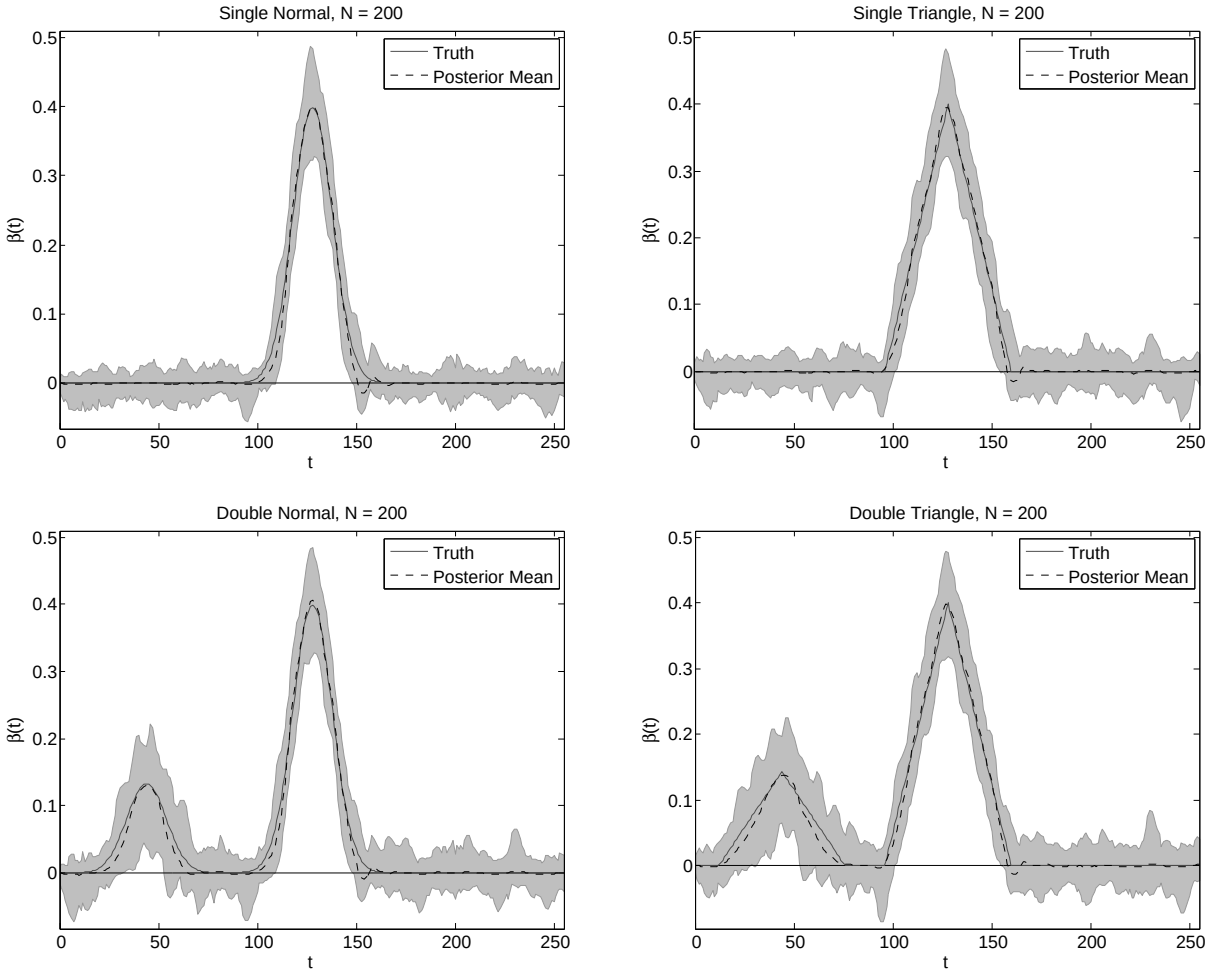


Figure 3.1: Posterior estimates of $\beta(t)$ as a function of t averaged over 200 simulated data sets for a single covariate. Light gray bands depict the 95th percentile across the simulated data sets, true functions are in solid dark gray. The top row contains estimates for the single peak scenarios, bottom row contains double peak.

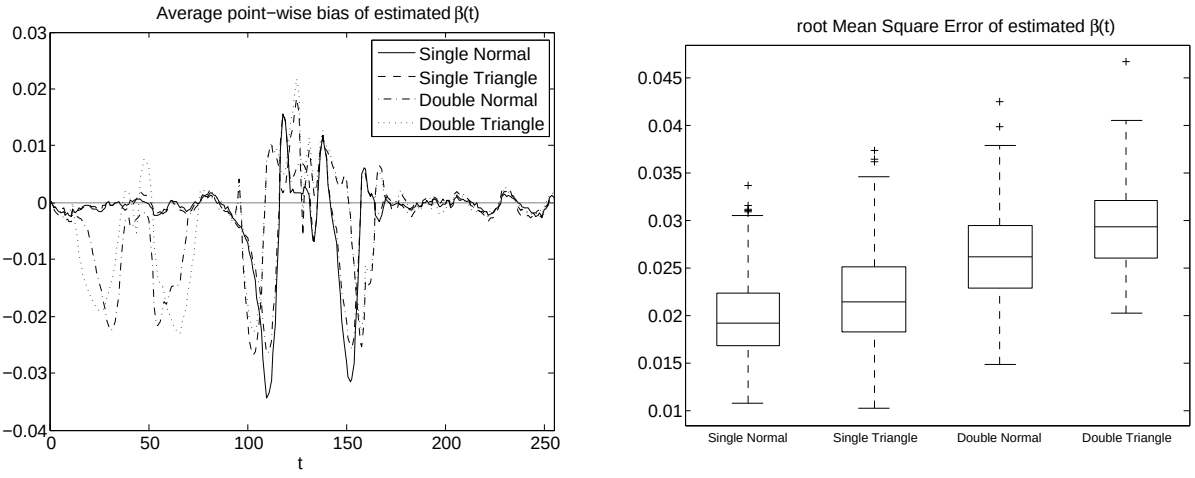


Figure 3.2: The figure on the left contains the average point-wise bias of the estimated $\beta(t)$ as a function of t taken across 200 simulated data sets. The figure on the right compares box plots across scenarios of rMSE calculated for each simulated data set.

rMSE values are similar across all four single covariates scenarios.

Figure 3.3 presents the results of the last scenario with two scalar covariates and increasing correlation. The left column contains the true values of $\beta_1(t)$ and $\beta_2(t)$ plotted along with the average estimated curve from each model. The gray bands in the left column are the 95th percentile of samples from the no correlation ($r = 0$) setting. From the left column, we see that when there is no correlation between predictors, the average estimates of $\beta_1(t)$ and $\beta_2(t)$ are similar to those from the single covariate case and effectively capture the truth. When correlation is moderate, $r = 0.5$, the model estimates $\beta_1(t)$ well but tends to over estimate the largest peak of $\beta_2(t)$. Not surprisingly, when correlation is high, $r = 0.9$, the model performs the worst though the estimates are still reasonably accurate.

This assessment is reinforced by the right column of Figure 3.3 which contains the average point-wise bias for each model. Here we see how similar the fits are when correlation is zero or moderate. In fact the moderate correlation has less bias around the peak than the no correlation setting. The high correlation setting has issues away from the peak when estimating $\beta_1(t)$ but has reasonable bias around the peak. In the more complicated

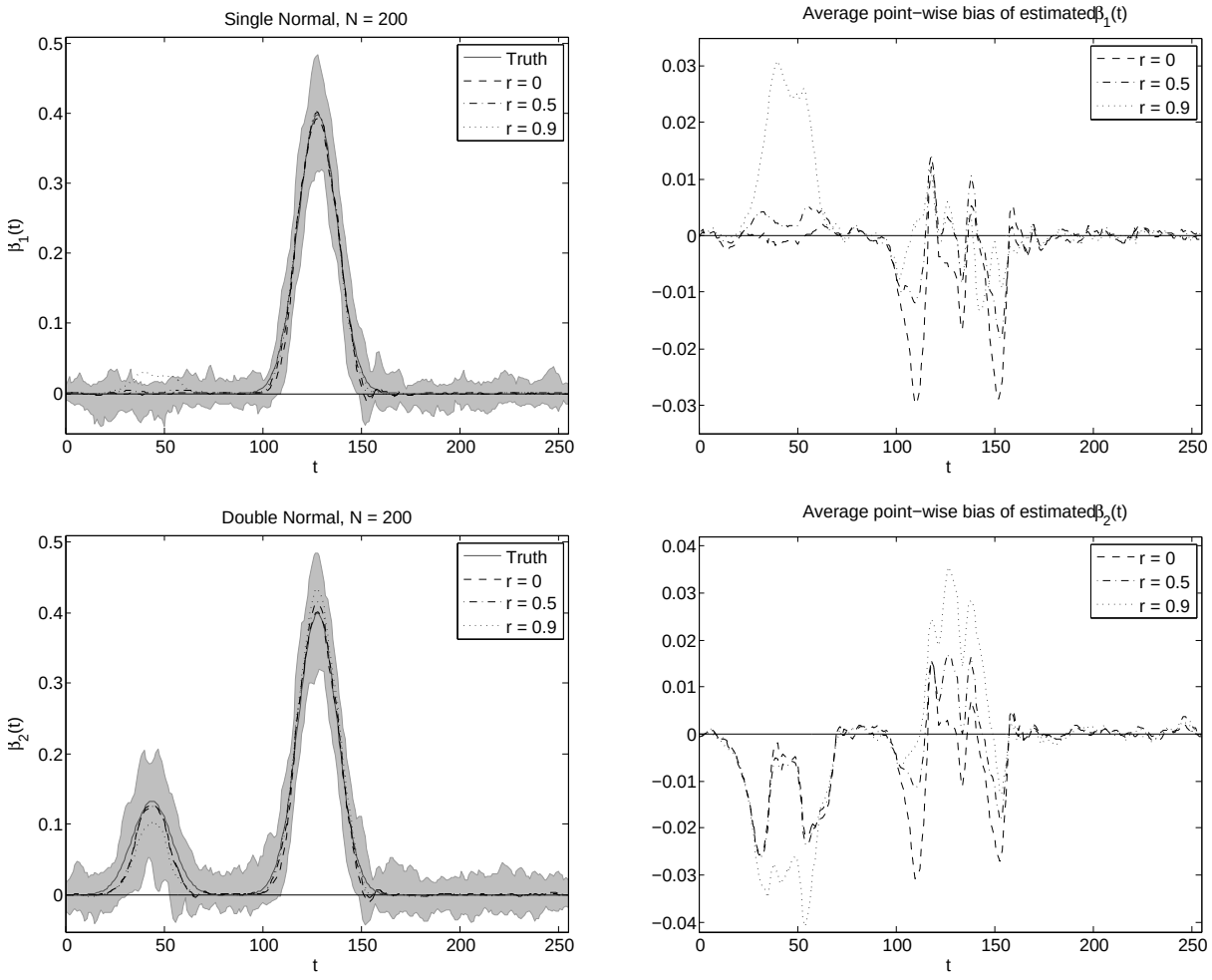


Figure 3.3: Posterior estimates of $\beta(t)$ as a function of t averaged over 200 simulated data sets for a single covariate. Light gray bands depict the 95th percentile across the simulated data sets, true functions are in solid dark gray. The top row contains estimates for the single peak scenarios, bottom row contains double peak.

double normal effect, again the no correlation and moderate correlation models perform similarly. This time the high correlation model underestimates the smaller peak while overestimating the larger peak. On the whole, this demonstrates that OPWAVFM is capable of handling multiple covariates with varying levels of correlation.

Figure 3.4 presents heat maps of both FDR_ε and SEN_Υ for the BFDR under varying levels of ε and Υ along the vertical axes, FDR_ε in the left column, SEN_Υ in the right. Since the single peak scenarios behaved similarly and both double peak scenarios also behaved similarly, we present only the results from the single normal and the double normal here. A similar figure to Figure 3.4 can be found in the Appendix. We see from Figure 3.4, that for values of π_δ near 0.5, FDR_ε tends to be high regardless of ε, though as π_δ increases, FDR goes to zero. FDR_ε values behave similarly across all four scenarios. For a given π_δ, SEN_Υ increases as Υ increases. At larger π_δ values BFDR does, not surprisingly, struggle with smaller Υ, the degree to which varies, however, by scenario. SEN_Υ for the single peak scenarios are still reasonably high even for large π_δ and small Υ. The double peak scenarios have low SEN_Υ even when both π_δ and Υ are in the middle of their ranges. This is in contrast to the single peaks which achieve reasonably high SEN_Υ rather quickly for all π_δ. Overall, the BFDR performs well in simulation in controlling FDR while detecting true signals.

As the SimBaS procedure does not require an intensity threshold, similar graphics to those in Figure 3.4 are not possible. Further, regardless of ε, under each scenario FDR_ε was zero. We do, however, display the values of SEN_Υ as functions of Υ for each scenario in Figure 3.5. Not surprisingly, as Υ increases, the SEN_Υ of SimBaS increases under each scenario. On the whole, the sensitivity of SimBaS is reasonable and comparable to that of the BFDR for larger π_δ. Interestingly, for small Υ, the single peak scenarios have larger SEN_Υ but for larger Υ, the double peak scenarios exhibit larger SEN_Υ.

Comparing the two inference procedures is difficult as the BFDR requires the selection of a π_δ. However, for large enough π_δ, false discovery rate is similar between BFDR and

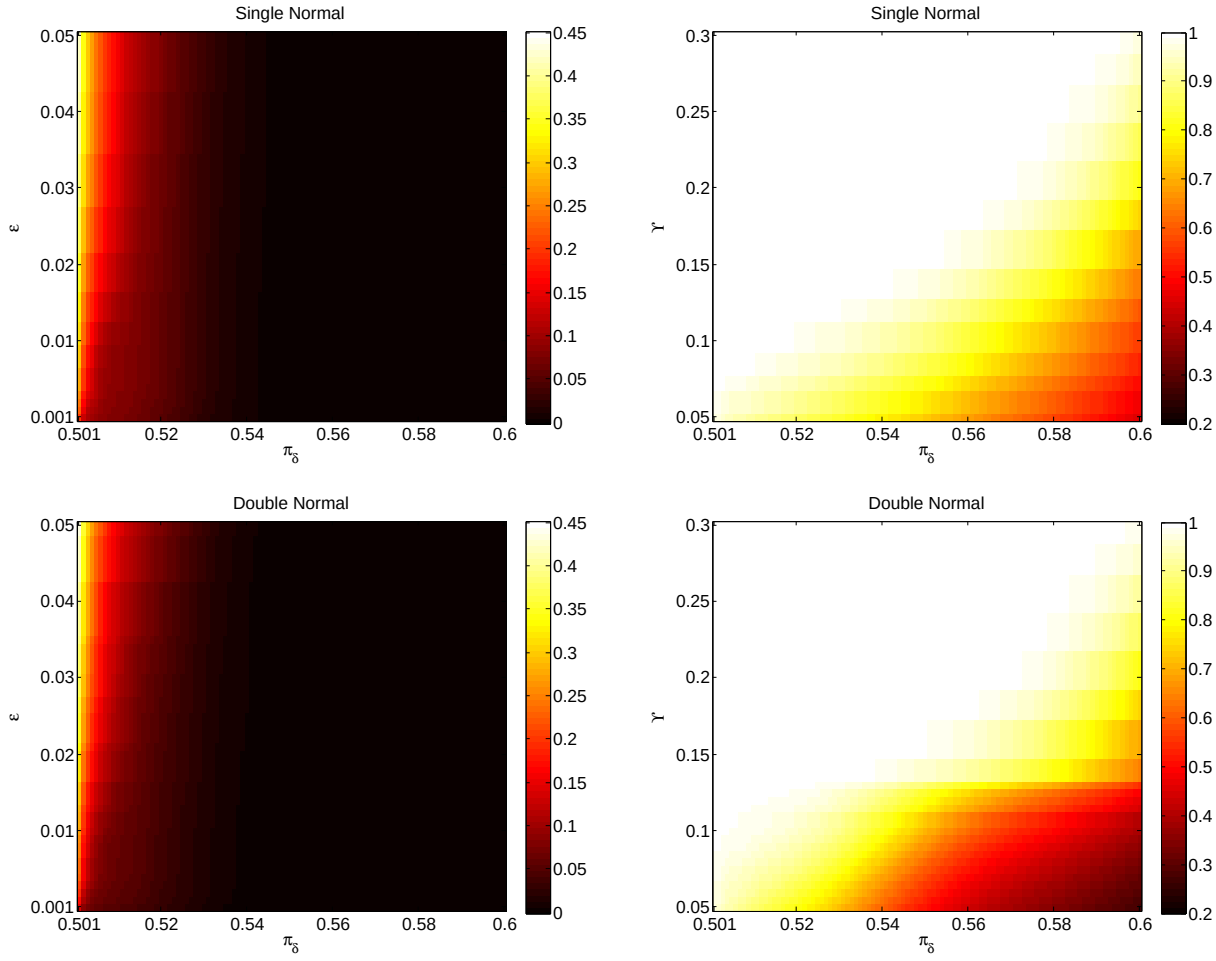


Figure 3.4: Left column: heat maps of FDR_ε as functions of ε for the BFDR with varying levels of π_δ. Right column: heat maps of SEN_Υ as functions of Υ for the BFDR with varying levels of π_δ

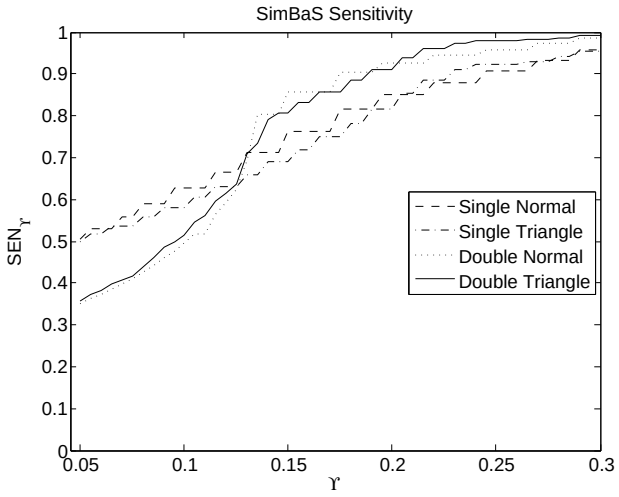


Figure 3.5: Plotted SEN$_\Upsilon$ as functions of Υ for the SimBaS procedure. Each single covariate scenario is depicted as described in the legend.

SimBaS across ε. Further, both procedures sensitivity increases as Υ increases. Sensitivity for SimBaS behaves similarly to the BFDR sensitivity for large π_δ. The tradeoffs between the two are apparent in that an appropriate choice of π_δ achieves a low false discovery rate while maintaing high sensitivity. However if an inappropriate π_δ is chosen, either FDR will be high or sensitivity will be low. SimBaS does not require a choice of threshold, maintains very low FDR and achieves high sensitivity for large Υ.

3.5 Application

Example data comes from a study of chronic obstructive pulmonary disease (COPD) which collected genotypes and expression profiles for 202 subjects. Of interest is genotyped SNPs from chromosome 15 near the IREB2 gene which has been previously identified as a disease susceptibility locus (Pillai et al., 2009; Wilk et al., 2009; Cho et al., 2010). Investigators also collected four different expression probe sets for IREB2 with probes starting at 78730518 bp and ending at 78793798 bp. We consider five models one for each probe set separately and one jointly modeling for the three probe sets that exhibited significant SNPs in single probe analysis. All models are adjusted for age, sex, array lot

number, and genetic ancestry. The four expression probe sets of interest are 1555476_at, 214666_x_at, 225892_at, and 242261_at.

Commonly, a fine mapping of SNPs around a candidate gene is sampled. One standard approach is to take all SNPs falling within a specified distance of the start and end of the gene, such as 250 kilobases (kb). Given the genotypes around IREB2 and taking 250kb to either side results in 190 total SNPs. Examining 2 megabases (Mb) to either side of the gene results in 1135 SNPs. As with the simulation, all application models were run using Daubechies wavelets with 4 vanishing moments and zero padding. Each model was run for 6000 total samples with the first 5000 discarded. Models taking 250kb to either side took just over an hour to finish while taking 2Mb to either side took just under three hours running on a high performance computing cluster. For brevity, only results from the 250kb models are presented here. Results from the 2Mb models can be found in the Appendix.

Since we're interested in identifying significantly associated SNPs, we further focus our analysis on first detecting SNPs that are significantly different from zero based on SimBaS. Next we examine those SNPs at varying intensity cutoffs using probabilities from the BFDR. Finally, we compare the association of these SNPs with different expression probe sets in both the single probe set models and the joint probe set model. Manhattan plots of SimBa Scores from the single probe set models can be found in Figure 3.6 while Figure 3.7 contains Scores from the joint probe set model. In each figure, the dotted-dashed gray line represents a global α of 0.05.

Ostensibly, the SimBa Score is a multiplicity adjusted probability testing the null hypothesis of no association between a specific SNP and the expression probe set. Thus as a first step, we can see in the single probe set models which SNPs are significantly associated with each probe set. From Figure 3.6, we see that all probe sets except 225892_at have significant SNPs in the single probe set model. Thus for our joint model, we focus in on the remaining three probe sets. To identify candidate SNPs, we use the joint model for the

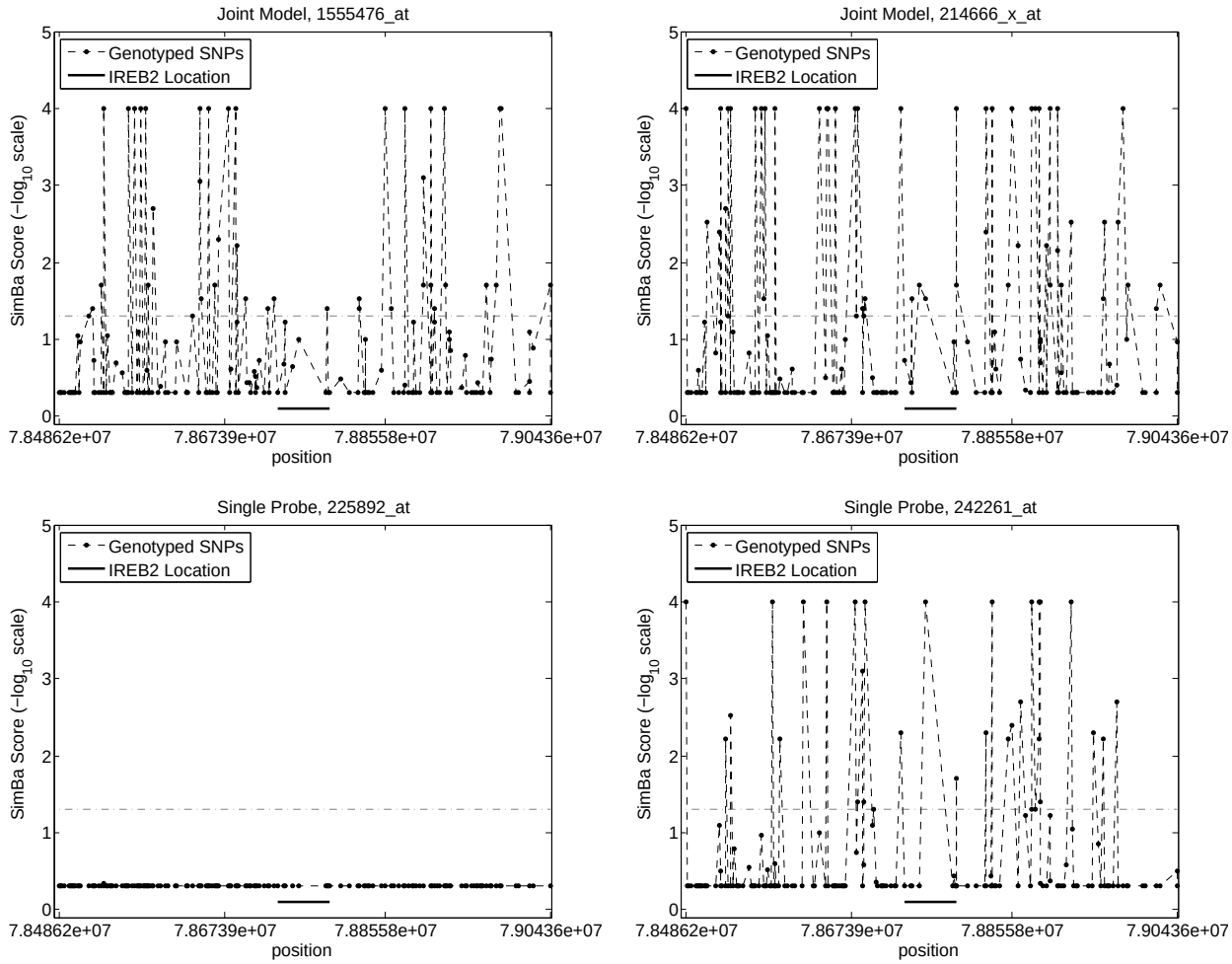


Figure 3.6: Single probe models SimBa Scores plotted as functions of position on the chromosome. The location of IREB2 is noted as a horizontal bar below the probabilities. For convenience, a dotted-dashed gray line depicts a global α-level of 0.05 plotted on the $-\log_{10}$ scale.

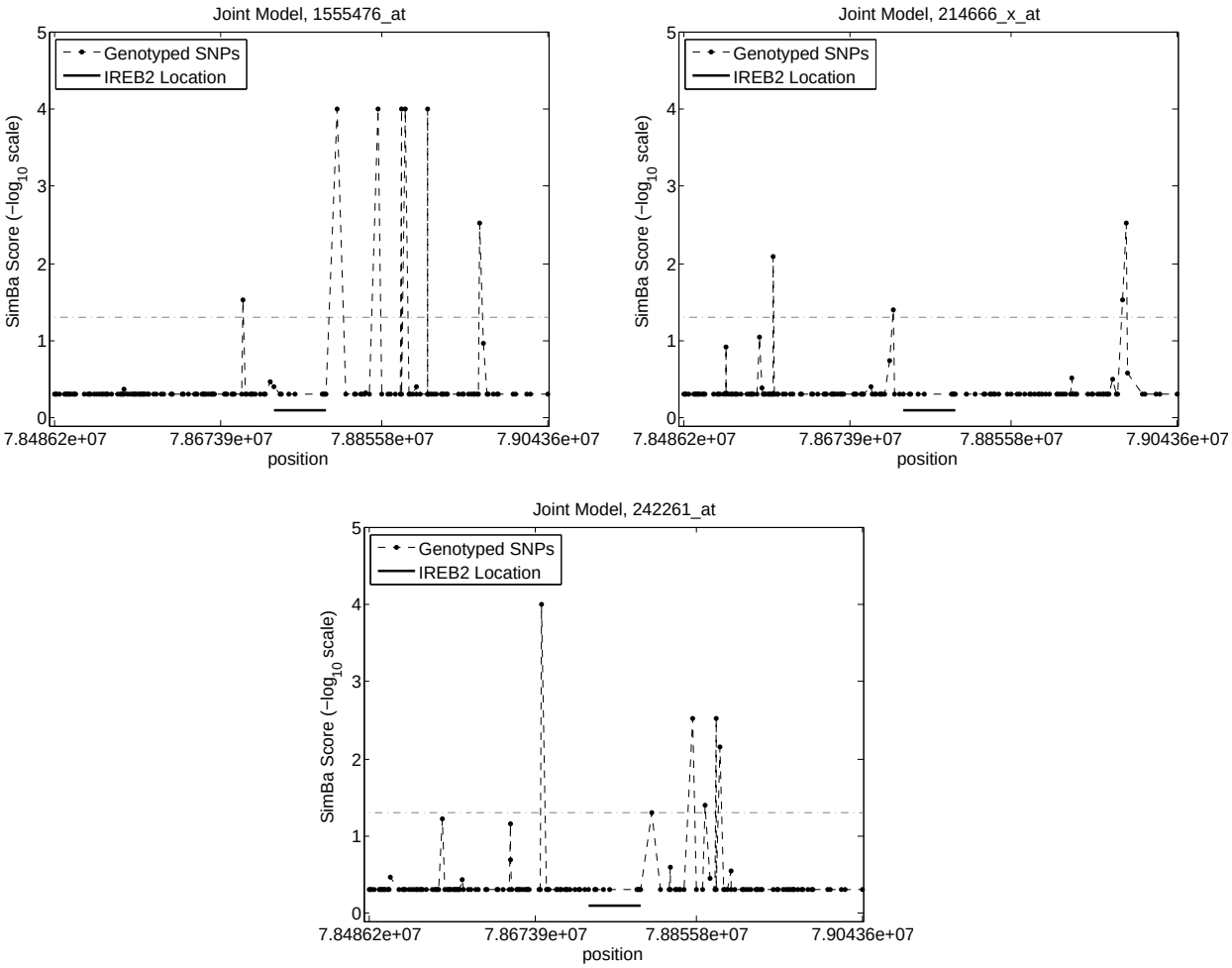


Figure 3.7: Joint model SimBa Scores plotted as functions of position on the chromosome. The location of IREB2 is noted as a horizontal bar below the probabilities. For convenience, a dotted-dashed gray line depicts a global α-level of 0.05 plotted on the $-\log_{10}$ scale.

Table 3.1: Significant SNPs from joint model by expression probe set. Significance based on joint model SimBa Score exceeding the global alpha of 0.05. Joint model scores are also compared to single probe scores for the same SNP as well as BFDR probabilities from the joint model for three different π_δ values. [1] denotes SNPs significantly associated with both 1555476_at and 242261_at.

			SimBaS		$P(\lvert\beta(t)\rvert > \pi_\delta)$		
Probe Set	SNP	Position	Joint	Single	0.54	0.56	0.58
1555476_at	rs7163013	78698759	3.00%	36.0%	99.6%	86.6%	55.9%
	rs8034191[1]	78806023	0.01%	33.0%	99.9%	99.9%	99.4%
	rs2036527[1]	78851615	0.01%	25.0%	99.9%	99.9%	99.9%
	rs951266	78878541	0.01%	39.0%	99.9%	99.9%	99.9%
	rs16969968[1]	78882925	0.01%	50.0%	99.9%	99.9%	98.6%
	rs6495308	78907656	0.01%	2.00%	99.0%	84.3%	17.9%
	rs11639372	78966655	0.30%	50.0%	80.0%	11.9%	0.10%
214666_x_at	rs1040262	78587910	0.80%	0.01%	14.4%	0.00%	0.00%
	rs2656057	78723082	4.00%	50.0%	98.1%	79.2%	39.0%
	rs2869550	78981001	3.00%	0.01%	52.1%	3.50%	0.00%
	rs3825806	78985342	0.30%	10.0%	70.6%	2.50%	0.00%
242261_at	rs4283201	78681365	0.01%	4.00%	91.3%	17.0%	0.10%
	rs8034191[1]	78806023	5.00%	50.0%	99.9%	99.1%	94.6%
	rs2036527[1]	78851615	0.30%	0.60%	99.9%	99.9%	99.0%
	rs6495306	78865893	4.00%	0.20%	99.9%	98.1%	88.8%
	rs951266	78878541	0.30%	5.00%	99.9%	99.9%	99.9%
	rs16969968[1]	78882925	0.70%	5.00%	99.9%	99.9%	95.7%

expression probe sets 1555476_at, 214666_x_at, and 242261_at considering only SNPs with SimBa Scores at or above the global α-level of 0.05. This results in seven SNPs associated with probe set 1555476_at, four with 214666_x_at, and six with 242261_at.

Table 3.1 contains the SimBa Scores and BFDR probabilities of SNPs detected in the joint the model. For the joint model, both SimBaS and BFDR were performed to each expression probe set separately. Likewise, SimBa Scores for the single probe model were calculated using only the coefficients of the expression probe set. Interestingly, for 1555476_at, only one of the SNPs identified was significant in the single probe set model while for 214666_x_at two of the four were significant in both models. This is in contrast to 242261_at where all but one SNP was found significant in both models.

Another interesting aspect comes from the BFDR. We can use the BFDR to both flag significant coefficients and to determine posterior probabilities that a give coefficient falls above a pre-determined threshold. For Table 3.1, we use three different thresholds on the probability scale: $\pi_\delta = 0.54, 0.56$, and 0.58. Examining all three allows us to see the magnitude at which these coefficients are significant. From the table we see that significance via the SimBa Score does not necessarily translate to a large effect size. The SimBa Score only tests is if the association is different from zero it does not give us an idea of the magnitude of effect. Of potential interest is determining SNPs that not only are significantly associated but also demonstrate meaningful effect sizes.

For instance, four of the SNPs that are significantly associated with probe set 1555476_at have BFDR probabilities above 90% for all values of π_δ. This suggests that not only are these SNPs significantly associated, but they also have a large magnitude of effect with over 0.90 probability of being larger than 0.58 on probability scale. Conversely, most of the SNPs that are significantly associated with probe set 214666_x_at do not have large effect sizes. One could consider using first using SimBaS to determine significantly associated SNPs and then selecting a sufficiently large cutoff π_δ to narrow in on SNPs that are not only significant but also have large magnitudes of effect.

One final observation is that each probe set is associated with different sets of SNPs. In fact there is no overlap between SNPs associated with 214666_x_at and the other two probe sets. However the remaining probe sets do have considerable overlap with three SNPs commonly associated with 1555476_at and 242261_at. All three of those SNPs, rs8034191, rs2036527, and rs16969968, maintain large BFDR probabilities as π_δ increases. SNPs rs2036527 and rs16969968 both have rather small SimBa Scores from the joint model while rs8034191's score falls right on the cutoff value.

3.6 Discussion

Functional data of all kinds, whether it be continuous or categorical, Gaussian or not, continues to be collected in vast quantities across many scientific disciplines. To keep up with the demand for methodologies and inferences for such data we must continue to expand the field of functional data analysis. While methods for Gaussian functional outcomes and scalar categorical outcomes with functional predictors are abundant in the literature, functional categorical outcomes have received little attention. While additional work is still needed, in extending the WFMM framework, we have developed a method capable of handling a functional categorical outcome along with a large set of predictors.

Here we've presented a novel model for performing function-on-scalar regression where the outcome comes from finely sampled categorical process. To implement this methodology, we've developed an MCMC procedure which builds on the framework of the WFMM while utilizing the latent variable representation of the probit model. As our eventual functional outcome of interest had three levels, we proposed an ordinal model assuming proportional probabilities which can easily be refined down to the binary case or extended to allow for more than three categories. Previous authors have shown the benefits of both the BFDR and SimBaS procedures (Meyer et al., 2014a,b), thus we reformulated both procedures for the OPWAVFM framework. Additionally, we propose an extension of the OPWAVFM to the function-on-function regression setting which also can be considered an extension of Meyer et al. (2014a) to generalized outcomes.

To evaluate the abilities of the OPWAVFM, we presented seven simulation scenarios which display the ability of our method to detect various true signals at sample sizes that are on the low end of genomic studies. Additionally, we examined the behavior of the model for multiple predictors of varying levels of correlation. In all scenarios, the OPWAVFM demonstrates low bias and rMSE. Even when correlation is induced between multiple covariates, bias is minimal. In fact, a moderate amount of correlation appears to improve the model fit. Also in simulation, we show that both the BFDR and SimBaS have

good properties displaying low false discovery rates and high sensitivity. And in application, we show how the OPWAVFM can be used to perform an eQTL analysis both with a single expression probe set of interest and with multiple. Further we note the ability of the SimBaS procedure to, in our modeling context, detect several significant SNPs as well as the ability of the BFDR to distinguish strength of association amongst those SNPs. While not presented here, the OPWAVFM is able to scale up to a larger number of SNPs at only a minimal computational cost.

One potential issue for further exploration is the frequency of measurement of the outcome in application. In particular, the wavelet decomposition we used assumes equally spaced measurements, an assumption which may be tenuous with SNP data. Further, the OPWAVFM is a proportional probability regression where the probability of being in one outcome category versus the next does not depend on the current category. Future work on the OPWAVFM will include exploration of the work done by Sardy et al. (1999) and Wand and Omerod (2011) on wavelet-based techniques for unequally spaced data. Additionally, Dunson and Neelon (2003) and Lin and Dunson (2013) proposed Bayesian procedures for order-constrained parameters which we hope to adapt for the OPWAVFM framework to allow for non-proportionality.

Acknowledgements

This work was supported by grants from the National Institutes of Health (ES007142, ES000002, ES016454, CA134294).

References

Cardot, H. and Sarda, P. (2005). Estimation in generalized linear models for functional data via penalized likelihood. *Journal of Multivariate Analysis* **92,** 24–41.

Castaldi, P. J., Cho, M. H., Qui, W., McGeachie, M., Celli, B., et al. (2014). Genetic Control of Gene Expression at Novel and Established Chronic Obstructive Pulmonary Disease Loci. *In Preparation*.

Cho, M. H., Boutaoui, N., Klanderman, B. J., Sylvia, J. S., Ziniti, J. P., et al. (2010). Variants in FAM13A are associated with chronic obstructive pulmonary disease. *Nature Genetics* **42,** 200–202.

Dunson, D. B. and Neelon, B. (2003). Bayesian Inference on Order-Constrained Parameters in Generalized Linear Models. *Biometrics* **59,** 286–295.

Flutre, T., Wen, X., Pritchard, X., and Stephens, M. (2013). A Statistical Framework for Joint eQTL Analysis in Multiple Tissues. *PLOS Genetics* **9**.

Frommlet, F., Ruhaltinger, F., Twaróg, P., and Bogdan, M. (2012). Modified versions of Bayesian Information Criterion for genome-wide association studies. *Computational Statistics and Data Analysis* **56,** 1038–1051.

James, G. M. (2002). Generalized linear models with functional predictors. *Journal of the Royal Statistical Society, Series B* **64,** 411–432.

Li, Y., Wang, N., and Carroll, R. J. (2010). Generalized Functional Linear Models with

Semiparametric Single-index Interactions. *Journal of the American Statistical Association* **105,** 621–633.

Lin, L. and Dunson, D. B. (2013). Bayesian Monotone Regression Using Gaussian Process Projection. arXiv:1306.4041v1.

Malloy, E. J., Morris, J. S., Adar, S. D., Suh, H., Gold, D. R., and Coull, B. A. (2010). Wavelet-based functional linear mixed models: an application to measurement error-corrected distributed lag models. *Biostatistics* **11,** 432–452.

McLean, M. W., Hooker, G., Staicu, A.-M., Scheipl, F., and Ruppert, D. (2014). Functional Generalized Additive Models. *Journal of Computational and Graphical Statistics* **23,** 249–269.

Meyer, M. J., Coull, B. A., Versace, F., Cinciripini, P., and Morris, J. S. (2014a). Bayesian Function-on-Function Regression for Multi-Level Functional Data. *In preparation*.

Meyer, M. J., Malloy, E. J., Morris, J. S., and Coull, B. A. (2014b). Bayesian Historical Functional Mixed Models for Repeated Measures. *In preparation*.

Morris, J. S., Brown, P. J., Herrick, R. C., Baggerly, K. A., and Coombes, K. R. (2008). Bayesian Analysis of Mass Spectrometry Proteomic Data Using Wavelet-Based Functional Mixed Models. *Biometrics* **64,** 479–489.

Morris, J. S. and Carroll, R. J. (2006). Wavelet-based functional mixed models. *Journal of the Royal Statistical Society, Series B* **68,** 179–199.

Müller, H.-G. and Stadtmüller, U. (2005). Generalized Functional Linear Models. *The Annals of Statistics* **33,** 774–805.

Müller, P., Parmigiani, G., and Rice, K. (2006). FDR and Bayesian multiple comparisons rules. *Johns Hopkins University, Dept. of Biostatistics Working Papers*.

Pillai, S. G., Ge, D., Zhu, G., Kong, X., Shianna, K. V., et al. (2009) A genome-wide association study in chronic obstructive pulmonary disease (COPD): identification of two major susceptibility loci. *PLoS Genet* **5:** e1000421.

Qiu W., Cho M. H., Riley J. H., Anderson W. H., Singh D., et al. (2011) Genetics of Sputum Gene Expression in Chronic Obstructive Pulmonary Disease. *PLoS ONE* **6(9)**: e24395. doi:10.1371/journal.pone.0024395

Ruppert, D., Wand, M. P., and Carroll, R. J. (2003). *Semiparametric Regression*. Cambridge University Press.

Sardy, S., Percival, D. B., Bruce, A. G., Gao, H.-Y., and Stuetzle, W. (1999). Wavelet shrinkage for unequally spaced data. *Statistics and Computing* **9,** 65–77.

Wand, M. P. and Ormerod, J. T. (2011). Penalized wavelets: Embedding wavelets into semiparametric regression. *Electronic Journal of Statistics* **5,** 1654–1717.

Wilk, J. B., Chen, T. H., Gottlieb, D. J., Walter, R. E., Nagle, M. W., et al. (2009) A genome-wide association study of pulmonary function measures in the Framingham Heart Study. *PLoS Genet* **5:** e1000429.

Żak-Szatkowska, M., and Bogdan, M. (2011). Modified versions of the Bayesian Information Criterion for sparse Generalized Linear Models. *Computational Statistics and Data Analysis* **55,** 2908–2924.

Zhu H., Brown, P. J., and Morris, J. S. (2011). Robust, adaptive functional regression in functional mixed model framework. *Journal of the American Statistical Association* **106,** 1167–1179.

Zhu H., Brown, P. J., and Morris, J. S. (2012). Robust classification of functional and quantitative image data using functional mixed models. *Biometrics* **68(4),** 1260–1268.

A. Appendices

A.1 Bayesian Function-on-Function Regression for Multi-Level Functional Data

A.1.1 MCMC Sampler

Working from Model (5) of the manuscript, the independence of the basis space allows us to split the model into T^* separate models for each basis coefficient in the y-space giving the model

$$\mathbf{y}^*_{(j,k)} = \mathbb{X}^* \beta^*_{(j,k)} + \mathbb{Z}\mathbf{u}^*_{(j,k)} + \mathbf{e}^*_{(j,k)}, \tag{3.1}$$

which is the same as Model (6) of the manuscript. We now place priors on the coefficients by noting that $\beta^*_{(j,k)} = \left\{\beta^*_{(p,jk)}\right\}$ where p indexes the V^{svd} retained principal components, $p = 1, \ldots, V^{svd}$. We place spike-and-slab priors on the elements of $\beta^*_{(j,k)}$ for a given j, k, and p via

$$\beta^*_{(p,jk)} \sim \ \gamma_{(p,jk)}\mathcal{N}(0, \tau_{pj}) + (1 - \gamma_{p,jk})d_0, \ \ \gamma_{(p,jk)} \sim \mathcal{B}(\pi_{pj})$$

where $\mathcal{B}$ denotes a Bernoulli distribution and d_0 represents a point-mass distribution at zero. Regularization parameters τ_{pj} and π_{pj} can be estimated using an Empirical Bayes-type approach as seen in Morris and Carroll (2006) and Malloy et al. (2010). Alternatively, priors may be placed as done in Zhu, Brown, and Morris (2011). For our model, we place an inverse gamma prior on the variances, τ_{pj}, of the Normal components of the mixture and a beta distribution on the mixture probabilities, π_{pj}, of the Bernoulli with respective hyper-parameters a_τ, b_τ, a_π, and b_π.

Following from Morris and Carroll (2006), we integrate out the random effects and work with marginalized likelihood. Morris and Carroll (2006) notes that this improves mixing over a naïve Gibbs sampler. The sampler alternates between sampling $\beta^*_{(j,k)}$ and the covariance parameters which we denote as $\mathbf{\Sigma}^*$. The random effects $\mathbf{u}^*_{(j,k)}$ are sampled when desired. The procedure iterates through the following steps:

Step 1: For each y-space coefficient indexed by (jk), we sample the fixed effect p from

the distribution $f(\beta^*_{(p,jk)}|\mathbf{y}^*_{(j,k)}, \beta^*_{(-p),jk}, \mathbf{\Sigma}^*)$ where $\beta^*_{(-p),jk}$ is the set of all fixed-effect coefficients at j, k except the pth. Morris and Carroll (2006) demonstrate that $f(\cdot)$ is a mixture of a point-mass at zero and a normal with mean $\mu_{p,jk}$ and variance $\varepsilon_{p,jk}$. The mixture probability $\alpha_{p,jk}$ is given by

$$\alpha_{p,jk} = \Pr\left(\gamma_{p,jk} = 1|\mathbf{y}^*_{(j,k)}, \beta^*_{(-p),jk}, \mathbf{\Sigma}^*\right) = O_{p,jk}/\left(O_{p,jk}+1\right)$$

where

$$O_{p,jk} = \pi_{pj}/(1-\pi_{pj})\text{BF}_{p,jk} \text{ and } \text{BF}_{p,jk} = (1+\tau_{p,jk}/V_{p,jk})^{-1/2}\exp\left\{\frac{1}{2}\zeta^2_{p,jk}(1+V_{p,jk}/\tau_{p,jk})\right\}$$

and the forms of the mean and variance of the normal are

$$\mu_{p,jk} = \hat{\beta}^*_{(p,jk),\text{MLE}}(1+V_{p,jk}/\tau_{p,jk})^{-1} \text{ and } \varepsilon_{p,jk} = V_{p,jk}(1+V_{p,jk}/\tau_{p,jk})^{-1}.$$

Step 2: For each y-space coefficient indexed by (jk), we next sample the elements $\sigma^2_{U(j,k)}$ and $\sigma^2_{E(j,k)}$ of Σ^*_U and Σ^*_E respectively. For this we use a random-walk Metropolis-Hastings step with objective function

$$\begin{aligned}&f(\sigma^2_{U(j,k)}, \sigma_{E(j,k)}|\mathbf{y}^*_{(j,k)}, \beta^*_{(j,k)}) \propto \\ &|\Sigma_{jk}|^{-1/2}\exp\left\{-\frac{1}{2}(\mathbf{y}^*_{(j,k)} - \mathbb{X}^*\beta^*_{(j,k)})'\Sigma^{-1}_{jk}(\mathbf{y}^*_{(j,k)} - \mathbb{X}^*\beta^*_{(j,k)})\right\}f(\sigma^2_{U(j,k)}, \sigma_{E(j,k)}).\end{aligned}$$

where Σ_{jk} is the marginal variance of $\mathbf{y}^*_{(j,k)}$. For the proposal distribution, we use an independent Gaussian truncated at zero and centered at the previous values.

Step 3: Random effects $U^*_{(j,k)}$ for each (j,k) are sampled from their full conditional which is a Gaussian distribution with mean $\{\Psi^{-1}_{jk} + 1/\sigma^2_{U(j,k)}\}^{-1}\Psi^{-1}_{jk}\hat{\mathbf{u}}_{NS,jk}$ and variance $\{\Psi^{-1} + 1/\sigma^2_{U(j,k)}\}^{-1}$, where $\Psi_{jk} = \{\mathbb{Z}'(1/\sigma^2_{E(j,k)})\mathbb{Z}\}^{-1}$ and

$$\hat{\mathbf{u}}_{NS,jk} = \left\{\mathbb{Z}'(1/\sigma^2_{E(j,k)})\mathbb{Z}\right\}^{-1}\mathbb{Z}'\left(1/\sigma^2_{E(j,k)}\right)\left(\mathbf{y}^*_{(j,k)} - \mathbb{X}^*\beta^*_{(j,k)}\right)$$

Step 4: Finally, we update τ_{pj} and π_{pj} separately from $f(\tau_{pj}|\gamma_{(p,jk)}, \beta^*_{(j,k)}, a_\tau, b_\tau)$ and $f(\pi_{pj}|\gamma_{(p,jk)}, a_\pi, b_\pi)$. The form of these conditionals are an inverse-gamma and beta respectively.

Notes: (1) Our approach can easily accommodate other shrinkage priors that might make sense for other basis functions, including Gaussians for spline bases, or other types of sparsity priors including Bayesian Lasso, Normal-Gamma, or Horseshoe Priors, which may have better sparsity and shrinkage properties under some settings; (2) These priors have connections to penalized likelihood methods, and their application in the basis space can induce smoothing or regularization across the coefficient surface $\boldsymbol{\beta}(v,t)$ in the data space; (3) The double-indexing inherent to multi-resolution bases like wavelets can be used for other bases, defining J clusters of basis coefficients $j = 1, \ldots, J$ containing K_j coefficients each, in order to allow clusters of coefficients to share common regularization parameters.

A.1.2 Additional Simulation Details

Equations for the four simulation scenarios discussed in the manuscript are found below:

$$\frac{7}{500}\frac{1}{\sqrt{(2\pi)0.003}}\exp\left[-\frac{1}{(2)(0.003)}\left(\frac{t}{225}-\frac{v}{225}\right)^2\right] \text{ (ridge)}, \tag{3.2}$$

$$\frac{437}{10000}\frac{1}{\sqrt{(2\pi)(0.03)}}\exp\left[-\frac{1}{(2)(0.03)}\left(\frac{t}{225}-\frac{v}{225}-0.5\right)^2\right] \text{ (lagged)}, \tag{3.3}$$

$$\frac{1029}{10000}\left[1+\frac{1}{1+\exp\left(\frac{0.25-\frac{t}{225}+\frac{v}{225}}{0.05}\right)}\right] \text{ (immediate)}, \tag{3.4}$$

$$\frac{1225}{10}(2\pi)^{-1}|\Sigma|^{-\frac{1}{2}}\exp\left[-\frac{1}{2}\left(\begin{bmatrix} v \\ t \end{bmatrix}-\begin{bmatrix} 150 \\ 60 \end{bmatrix}\right)'\Sigma^{-1}\left(\begin{bmatrix} v \\ t \end{bmatrix}-\begin{bmatrix} 150 \\ 60 \end{bmatrix}\right)\right] \text{ (peak)} \tag{3.5}$$

where

$$\Sigma = \begin{bmatrix} 15 & 0.5(15)(15) \\ 0.5(15)(15) & 15 \end{bmatrix}.$$

Graphical representations of each of these can be found both in the manuscript as well as in Figure A.1.

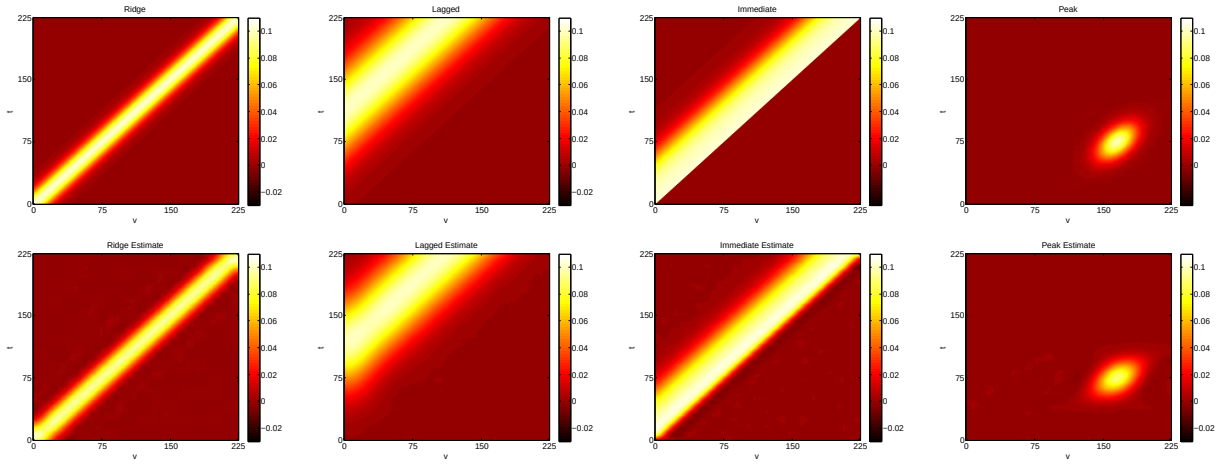


Figure A.1: Heat maps of proposed "true" surfaces for simulation study are in the top row, equations found in Models (3.2) - (3.5). The surfaces are estimated based on a sample size of $n = 100$ with repeated measures giving a total of $N = 200$ observations.

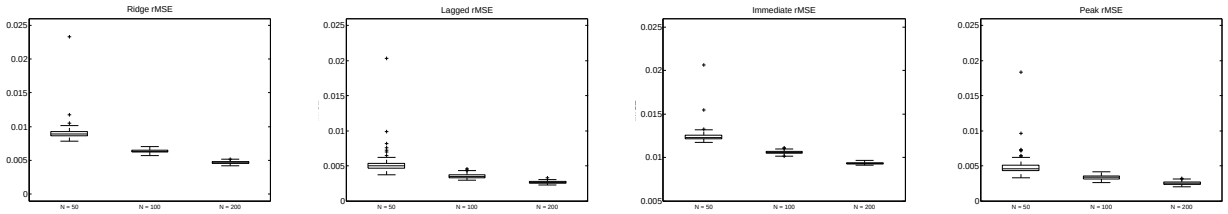


Figure A.2: Boxplots comparing the root-Mean Square Error amongst varying sample sizes by simulation scenario. Not surprisingly, as sample size increases, rMSE decreases. Note that the graphics are listed by total number of observations with $N = 50$ on the left and $N = 200$ on the right.

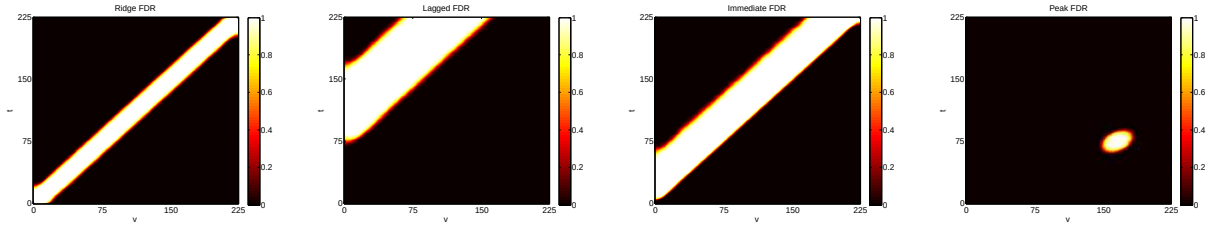


Figure A.3: Each figure contains a heat map displaying the FDR acceptance region averaged over 200 simulated data sets. The dark red regions indicate coefficients flagged in every or almost every data set. Dark blue coefficients were not flagged in any or almost any data sets. At the edge of each flagged region, coefficients that were flagged only occasionally can be seen. These figures were based on the smallest sample size $N = 50$, $n = 25$.

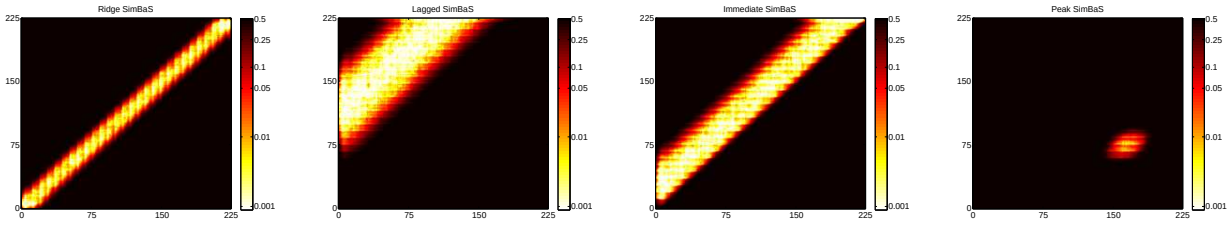


Figure A.4: Each figure contains a heat map displaying the SimBa Scores. All plots are on the log-scale, however for interpretability, the color scale has been exponentiated. For consistency of interpretation, the color scale was reversed so that the darker the red the more significant the coefficient and the darker the blue, the less significant. The scale is set to exhibit the variation in the SimBa Scores over the region of elevated significance. As such, the max of the scale does not accurately reflect the SimBa Scores where the true surface lacks association. Each plot represents the average SimBaS for each coefficient over the 200 simulated data sets. These figures were based on the smallest sample size $N = 50$, $n = 25$.

A.1.3 Additional Application Results

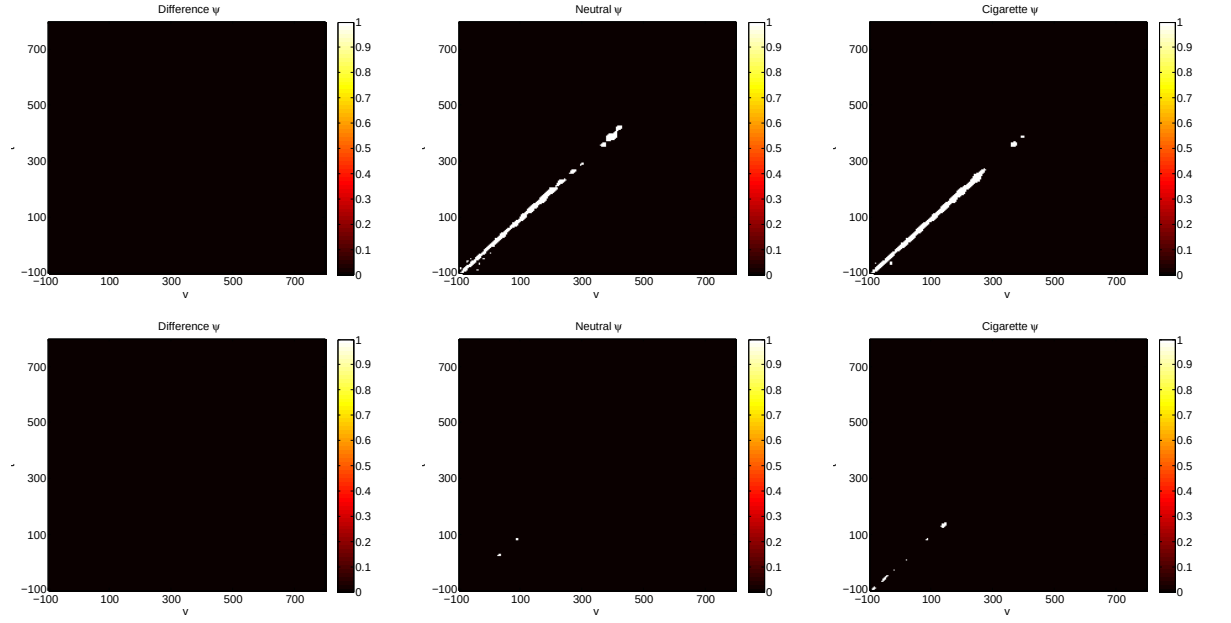


Figure A.5: Heat maps of the set of flagged locations (v, t), ψ, from the BFDR. Significant locations appear in white, non-significant locations are in black. For the difference surface, $\alpha = 0.05$ while the image-specific surfaces use $\alpha = 0.025$. The δ-intensity change is 0.05 for all surfaces. The top row contains results from the model using sensors 129 and 55. The bottom row contains results from the model using sensors 75 and 11.

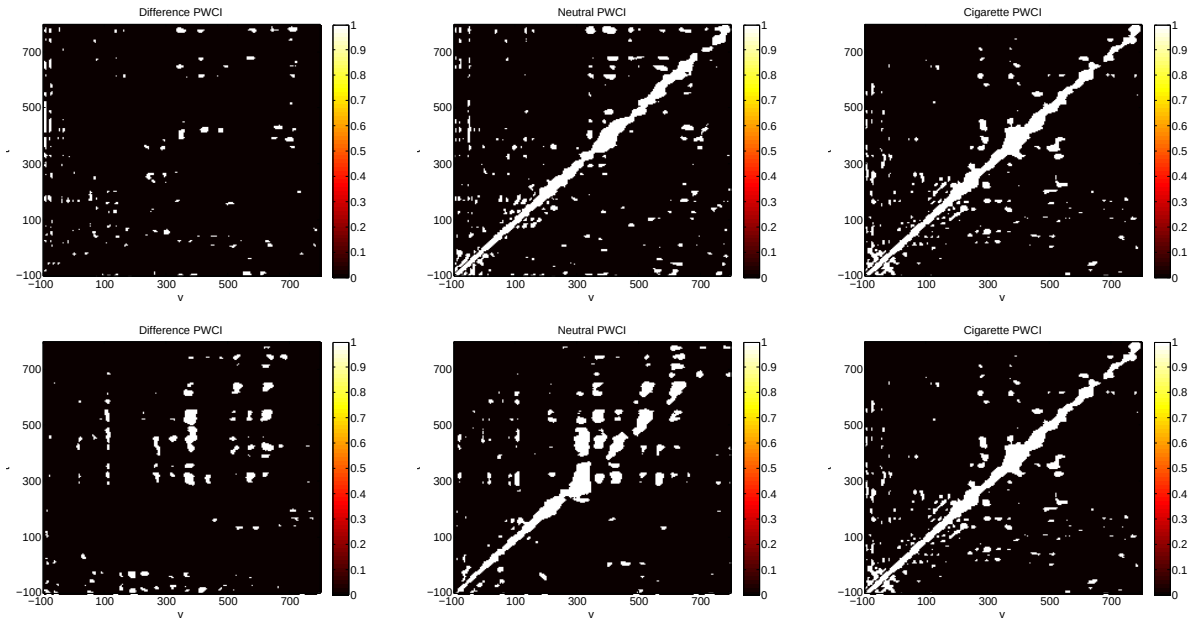


Figure A.6: Heat maps of the set of flagged locations (v, t) using PWCI. Locations in white were flagged as significant by the procedure while locations in black are not-significant. The top row contains results from the model using sensors 129 and 55. The bottom row contains results from the model using sensors 75 and 11.

A.2 Bayesian Historical Functional Mixed Models for Repeated Measures

A.2.1 MCMC Sampler

The MCMC sampler for our historical wavelet-packet model is derived from the sampler found in both Morris and Carroll (2006) and Meyer et al. (2014). The major difference from Meyer et al. (2014) here is that we only sample coefficients satisfying the constraint $\ell \leq k$. The prior we place on the coefficients satisfies this constraint and is

$$\beta^{W_P}_{(s\ell,jk)} \sim \ 1(\ell \leq k)\gamma_{(s\ell,jk)}\mathcal{N}(0,\tau_{jk}) + (1-\gamma_{s\ell,jk})d_0, \ \ \gamma_{(s\ell,jk)} \sim \mathcal{B}(\pi_{jk}).$$

From here, the sample follows in a similar manner to both Morris and Carroll (2006) and Meyer et al. (2014).

Step 1: Working by each y-space coefficient, sample the fixed effect at (jk) and $s\ell$ such that $\ell \leq k$ from the distribution $f(\beta^*_{(s\ell,jk)}|\mathbf{y}^*_{(j,k)}, \beta^*_{(-s\ell),jk}, \mathbf{\Sigma}^*)$ where $\beta^*_{(-s\ell),jk}$. $f(\cdot)$ is then a mixture of a point-mass at zero and a normal with mean $\mu_{s\ell,jk}$ and variance $\varepsilon_{s\ell,jk}$ with mixture probability $\alpha_{s\ell,jk}$ given by

$$\alpha_{s\ell,jk} = \Pr\left(\gamma_{s\ell,jk} = 1|\mathbf{y}^*_{(j,k)}, \beta^*_{(-s\ell),jk}, \mathbf{\Sigma}^*\right) = O_{s\ell,jk}/\left(O_{s\ell,jk}+1\right)$$

where

$$O_{s\ell,jk} = \pi_{s\ell j}/(1-\pi_{pj})\text{BF}_{s\ell,jk}$$

and

$$\text{BF}_{s\ell,jk} = (1+\tau_{s\ell,jk}/V_{s\ell,jk})^{-1/2}\exp\left\{\frac{1}{2}\zeta^2_{s\ell,jk}(1+V_{s\ell,jk}/\tau_{s\ell,jk})\right\}.$$

The mean and variance are then

$$\mu_{s\ell,jk} = \hat{\beta}^*_{(s\ell,jk),\text{MLE}}(1+V_{s\ell,jk}/\tau_{s\ell,jk})^{-1} \text{ and } \varepsilon_{s\ell,jk} = V_{s\ell,jk}(1+V_{s\ell,jk}/\tau_{s\ell,jk})^{-1}.$$

Step 2: Next sample the elements $\sigma^2_{U(j,k)}$ and $\sigma^2_{E(j,k)}$ of Σ^*_U and Σ^*_E respectively using a random-walk Metropolis-Hastings step with objective function

$$f(\sigma^2_{U(j,k)}, \sigma_{E(j,k)}|\mathbf{y}^*_{(j,k)}, \beta^*_{(j,k)}) \propto$$
$$|\Sigma_{jk}|^{-1/2} \exp\left\{-\frac{1}{2}(\mathbf{y}^*_{(j,k)} - \mathbb{X}^*\beta^*_{(j,k)})'\Sigma^{-1}_{jk}(\mathbf{y}^*_{(j,k)} - \mathbb{X}^*\beta^*_{(j,k)})\right\} f(\sigma^2_{U(j,k)}, \sigma_{E(j,k)}).$$

where Σ_{jk} is the marginal variance of $\mathbf{y}^*_{(j,k)}$. The proposal distribution is an independent Gaussian truncated at zero and centered at the previous values.

Step 3: Sample random effects $\mathbf{u}^*_{(j,k)}$ from their full conditional which is a Gaussian distribution with mean $\{\Psi^{-1}_{jk} + 1/\sigma^2_{U(j,k)}\}^{-1}\Psi^{-1}_{jk}\hat{\mathbf{u}}_{NS,jk}$ and variance $\{\Psi^{-1} + 1/\sigma^2_{U(j,k)}\}^{-1}$, where $\Psi_{jk} = \{\mathbb{Z}'(1/\sigma^2_{E(j,k)})\mathbb{Z}\}^{-1}$ and

$$\hat{\mathbf{u}}_{NS,jk} = \left\{\mathbb{Z}'(1/\sigma^2_{E(j,k)})\mathbb{Z}\right\}^{-1}\mathbb{Z}'\left(1/\sigma^2_{E(j,k)}\right)\left(\mathbf{y}^*_{(j,k)} - \mathbb{X}^*\beta^*_{(j,k)}\right).$$

Step 4: Finally, update τ_{pj} and π_{pj} separately using $f(\tau_{pj}|\gamma_{(p,jk)}, \beta^*_{(j,k)}, a_\tau, b_\tau)$ which is an inverse-gamma and $f(\pi_{pj}|\gamma_{(p,jk)}, a_\pi, b_\pi)$ which is beta.

A.2.2 Additional Simulation Details

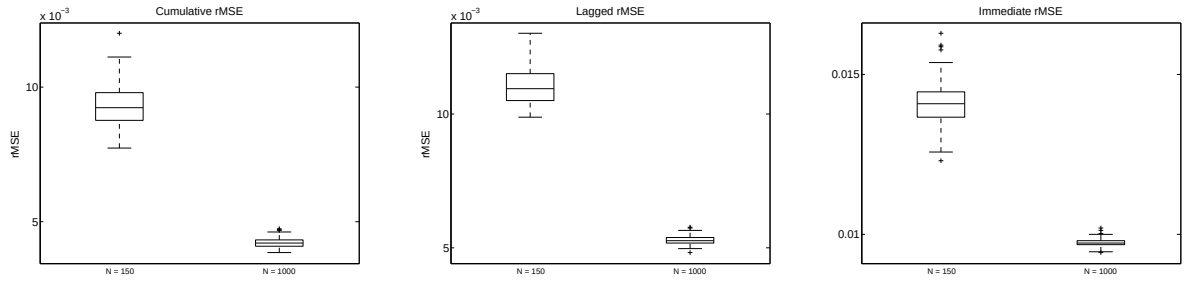


Figure A.7: Boxplots of rMSE by total sample size. For $N = 150$, there were only $n = 45$ subjects. For $N = 1000$, there were $n = 1000$ subjects.

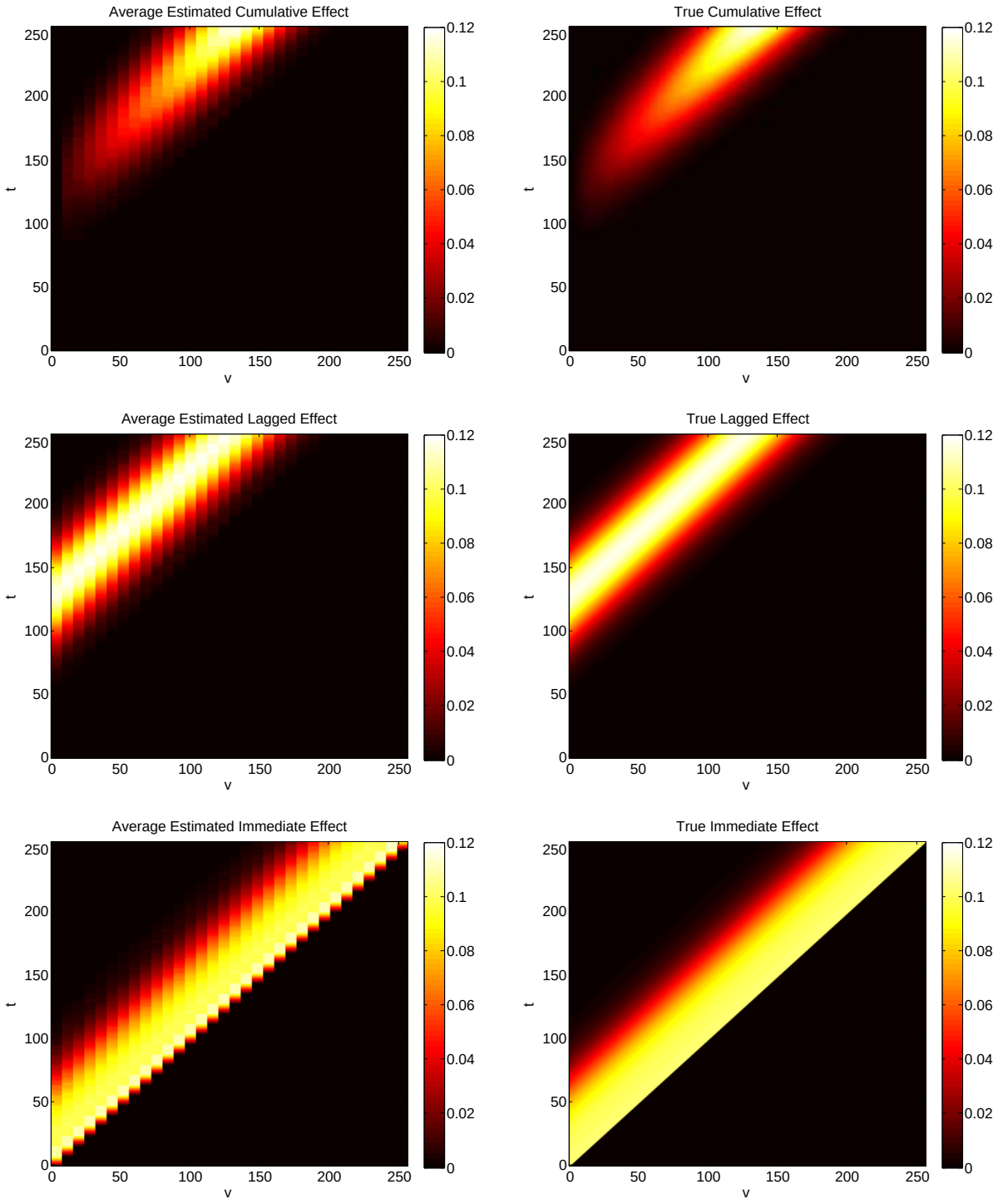


Figure A.8: Left column: heat maps of average estimated $\beta(v,t)$ plotted as functions of t and v based on a sample size of $N = n = 1000$. Right column: heat maps of the true $\beta(v,t)$ functions plotted as functions of t and v.

A.3 Ordinal Probit Wavelet-based Functional Models for eQTL Analysis

A.3.1 Additional Simulation Details

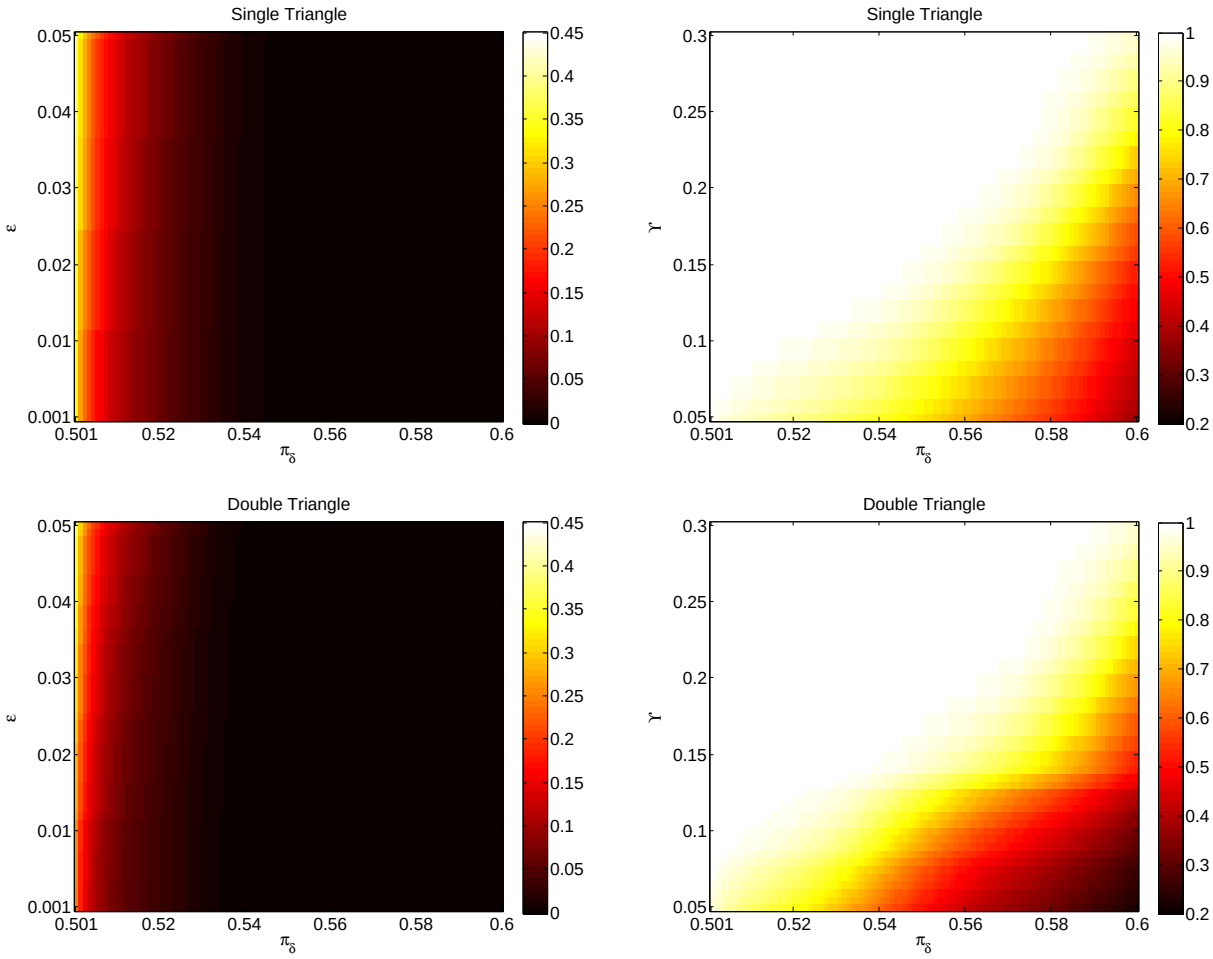


Figure A.9: Left column: heat maps of FDR_ε as functions of ε for the BFDR with varying levels of π_δ. Right column: heat maps of SEN_Υ as functions of Υ for the BFDR with varying levels of π_δ

A.3.2 Additional Results for SNPs within 250kb of IREB2

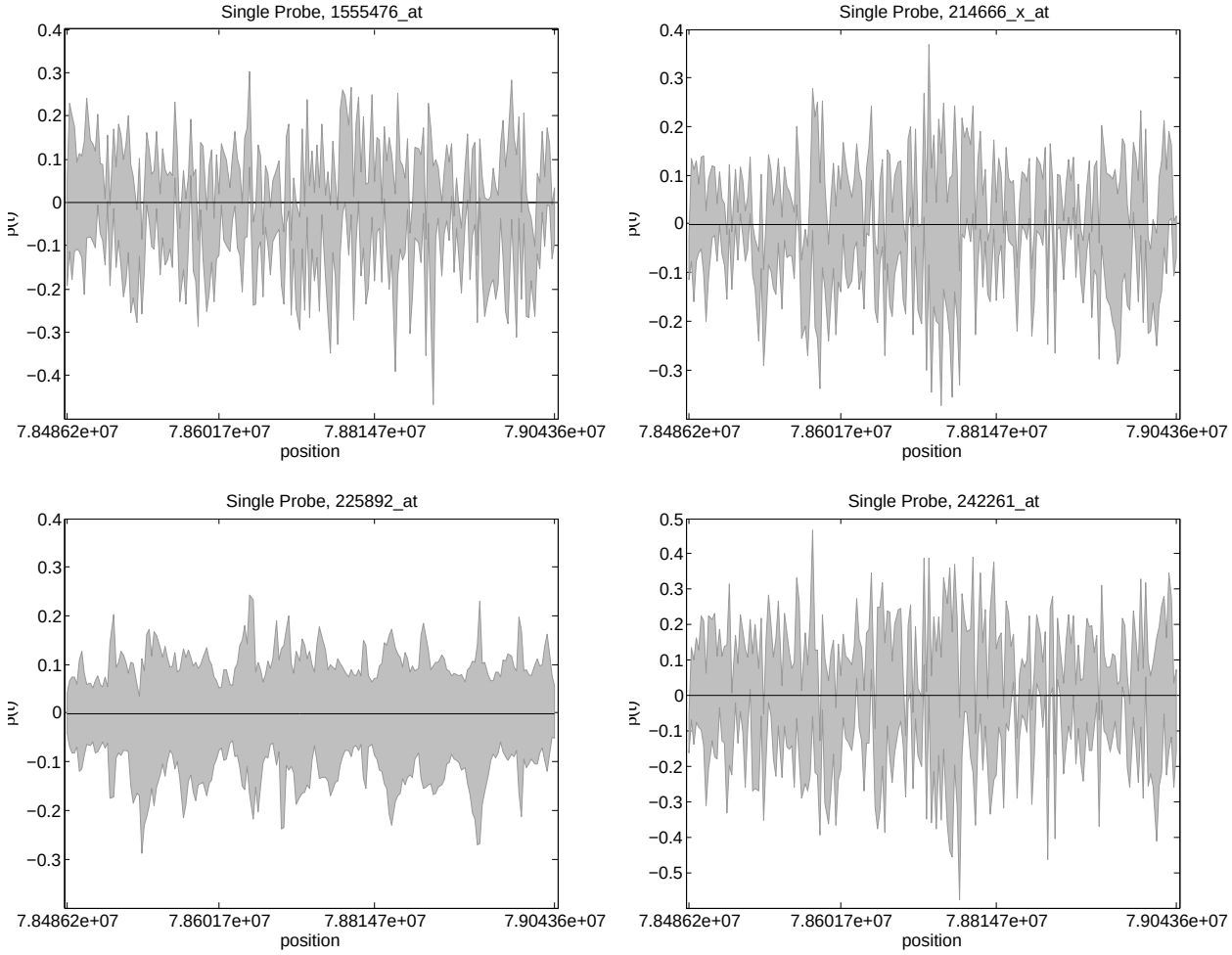


Figure A.10: Joint 95% credible bands for each single probe set model. Bands are calculated in the manner describe both in the paper and in Ruppert, Wand, and Carroll (2003).

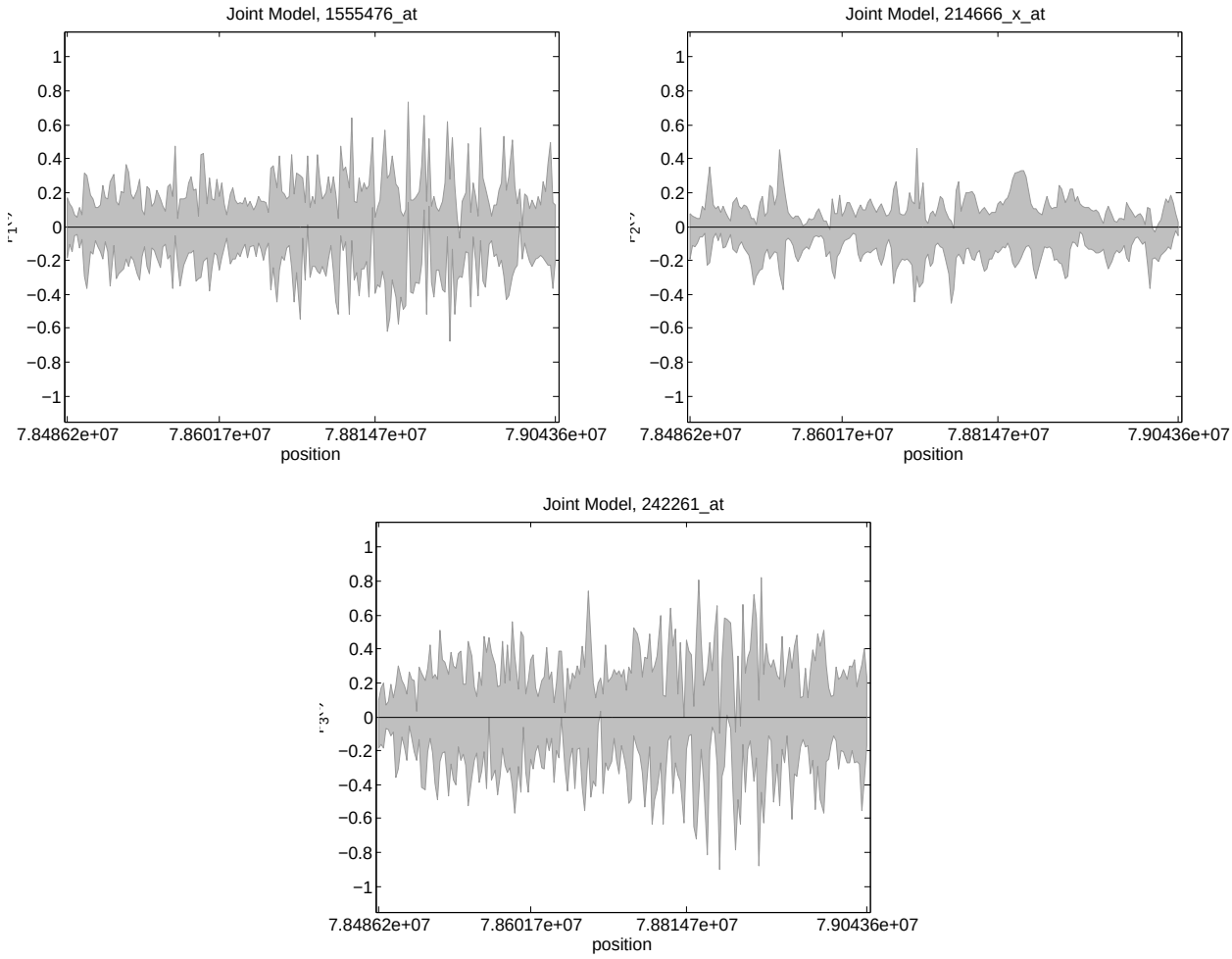


Figure A.11: Joint 95% credible bands for each probe set from the joint model. Bands are calculated on each probe set separately in the manner describe both in the paper and in Ruppert, Wand, and Carroll (2003).

A.3.3 Additional Results for SNPs within 2Mb of IREB2

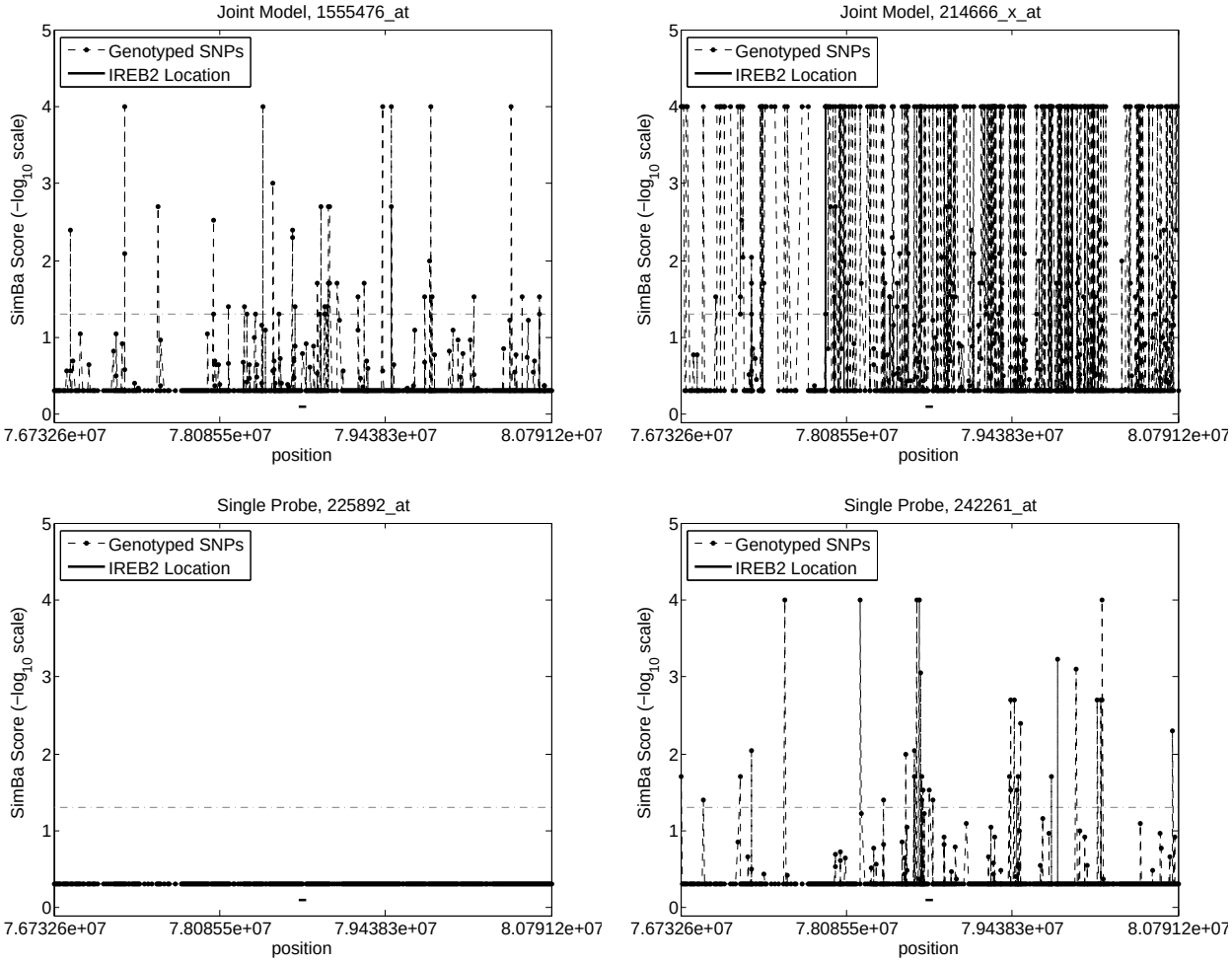


Figure A.12: Single probe models SimBa Scores plotted as functions of position on the chromosome. The location of IREB2 is noted as a horizontal bar below the probabilities. For convenience, a dotted-dashed gray line depicts a global α-level of 0.05 plotted on the $-\log_{10}$ scale. Scores are from the models taking 2Mb to either side of IREB2.

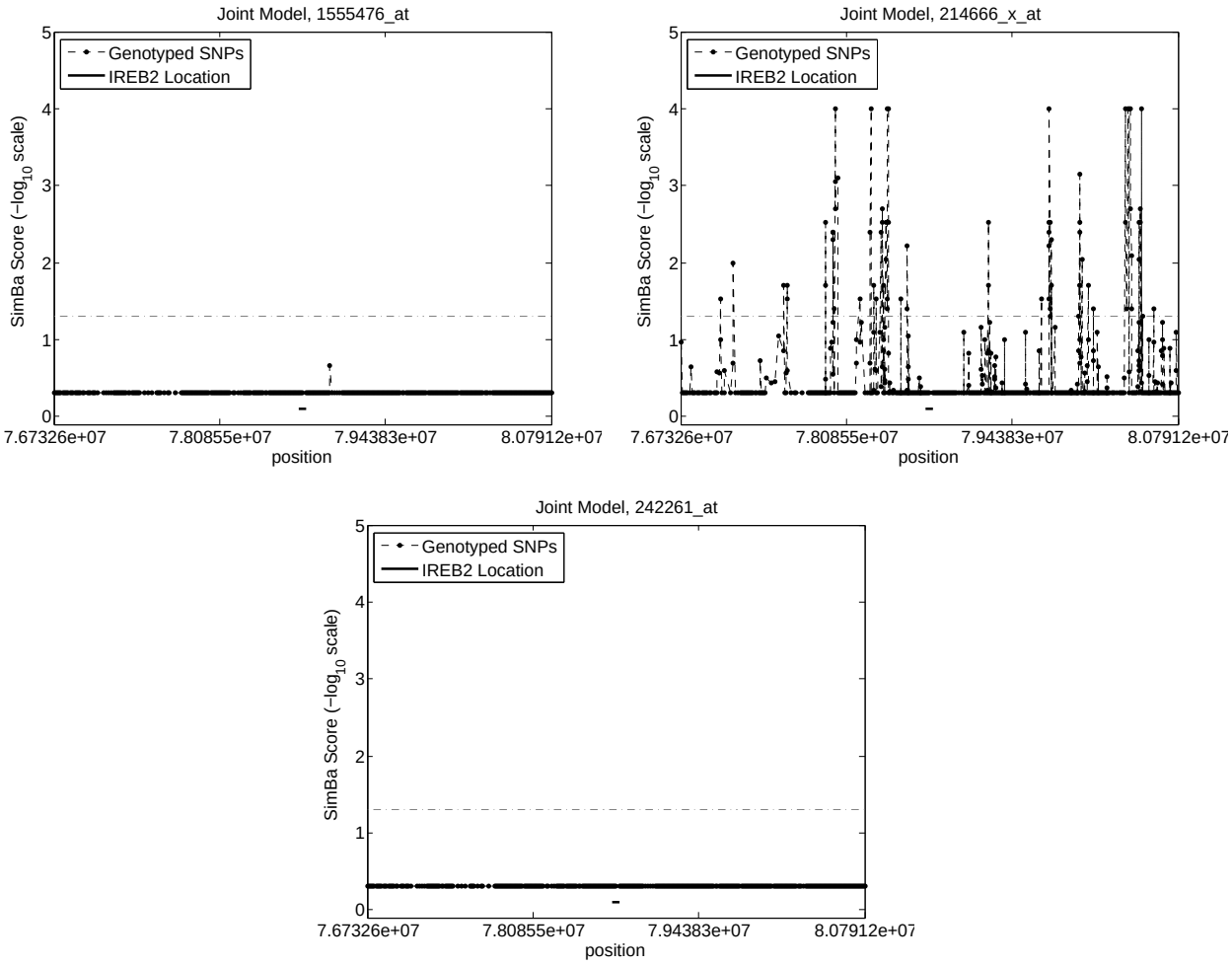


Figure A.13: Joint model SimBa Scores for the probe sets 1555476_at, 214666_x_at, and 242261_at plotted as functions of position on the chromosome. The location of IREB2 is noted as a horizontal bar below the probabilities. For convenience, a dotted-dashed gray line depicts a global α-level of 0.05 plotted on the $-\log_{10}$ scale. Scores are from the models taking 2Mb to either side of IREB2.

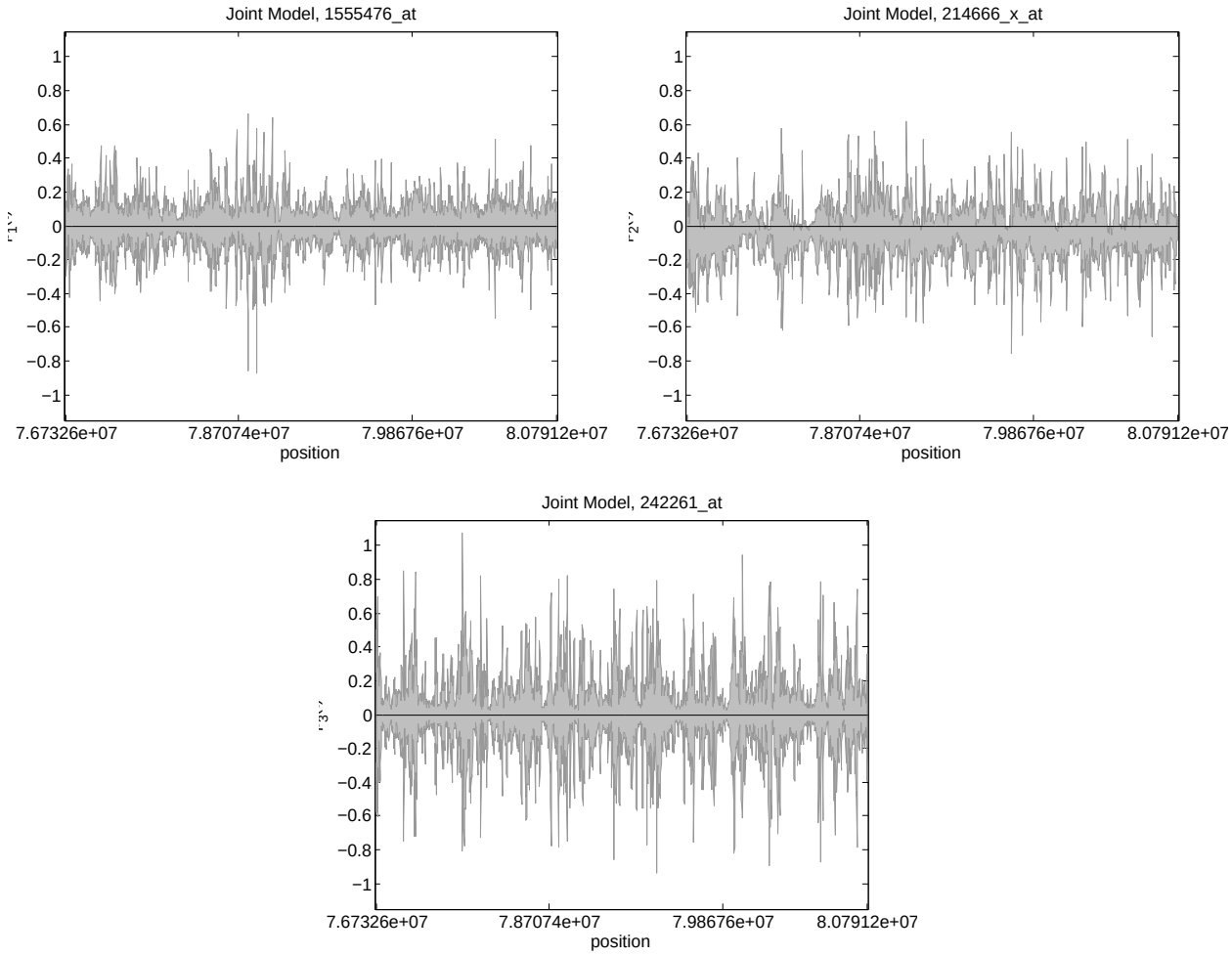


Figure A.14: Joint 95% credible bands for each probe set from the joint model for 1555476_at, 214666_x_at, and 242261_at. Bands are calculated on each probe set separately in the manner describe both in the paper and in Ruppert, Wand, and Carroll (2003).

References

Malloy, E. J., Morris, J. S., Adar, S. D., Suh, H., Gold, D. R., and Coull, B. A. (2010). Wavelet-based functional linear mixed models: an application to measurement error-corrected distributed lag models. *Biostatistics* **11,** 432–452.

Meyer, M. J., Coull, B. A., Versace, F., and Morris, J. S. (2014). Bayesian Function-on-Function Regression for Multi-Level Functional Data. *In Preparation*.

Morris, J. S. and Carroll, R. J. (2006). Wavelet-based functional mixed models. *Journal of the Royal Statistical Society, Series B* **68,** 179–199.

Ruppert, D., Wand, M. P., and Carroll, R. J. (2003). *Semiparametric Regression*. Cambridge University Press.

Zhu H., Brown, P. J., and Morris, J. S. (2011). Robust, adaptive functional regression in functional mixed model framework. *Journal of the American Statistical Association* **106,** 1167–1179.